SID® NUCLEIC ACIDS AND POLYPEPTIDES SELECTED FROM A PATHOGENIC STRAIN OF HEPATITIS C VIRUS AND APPLICATIONS THEREOF

FIELD OF THE INVENTION

The present invention relates to nucleic acids encoding SID® polypeptides which bind selectively to a polypeptide encoded by a pathogenic strain of the hepatitis C virus, as well as to the SID® polypeptides which are encoded by said nucleic acids.

The invention also concerns vectors comprising a nucleic acid encoding a SID® polypeptide as well as host cells transformed with such vectors.

The invention is also directed to two-hybrid methods which make use of the nucleic acids encoding a SID® polypeptide selected from a pathogenic strain of the hepatitis C virus as well as to methods for selecting molecules which inhibit the binding between a SID® polypeptide and a polypeptide which specifically binds thereto.

The invention also pertains to marker compounds containing a SID® polypeptide as well as nucleic acids encoding such marker compounds and methods and kits using the same.

BACKGROUND OF THE INVENTION

The hepatitis C virus (HCV) causes several liver diseases, including liver cancer. The HCV genome is a plus-stranded RNA that encodes the single polyprotein processed into at least 10 mature polypeptides.

The structural proteins are located in the amino terminal quarter of the polyprotein, and the non-structural (NS) polypeptides in the remainder (for a review, see HOUGHTON, 1996). The genome organisation resembles that of flaviviruses and pestiviruses and HCV is now considered to be a member of the *flaviviridae* family.

The gene products of HCV are, from the N-terminus to the C-terminus: core (p22), E1 (gp35), E2 (gp70), NS2(p21), NS3 (p70), NS4a (p4), NS4b(p27), NS5a (p58), NS5b (p66), as disclosed in figure 1. Core, E1 and E2 are the structural proteins of the virus processed by the host signal peptidase(s). The core protein and the genomic RNA constitute the internal viral core and E1 and E2 together with lipid membrane constitute the viral envelop (DUBUISSON et al., 1994; GRAKOUI et al., 1993; HIGIKATA et al. , 1993.).

The NS proteins are processed by the viral protein NS3 which has two functional domains: one (Cro-1), encompassing the NS2 region and the N-terminal portion of NS3, which cleaves autocatalytically between NS2 and NS3, and the other (Cro-2), located solely in the N-terminal portion of NS3, cleaves the other sites downstream NS3 (BARTENSCHLAGER et al; 1995; HIGIKATA et al;, 1993).

Various HCV protein-protein interactions have already been identified, notably by two hybrid methods. Noticeably, FLAJOLET et al; (2000) have shown interactions between NS3 and NS4A proteins as well as between NS4A and NS2 proteins. These authors have also shown core-core, NS3-E2, NS5A-E1, NS4A-NS3 and NS4A-NS2 interactions. Covalent as well as non-covalent interactions between E1 and E2 have been shown by PATEL et al; (1999). The protein interactions between NS3 and the HCV RNA helicase have also been described (MIN et al; 1999; GALLINARI et al., 1999) as well as interaction between NS3 and NS4A (URBANI et al. , 1999; DI MARCO et al., 2000; BUTKIEWICZ et al. , 2000).

However, the prior art methods allow the determination of interactions between full length proteins or large domains of proteins encoded by the genome of the hepatitis C virus which may contain more than one region of interaction with one or several HCV proteins. BUTKIEWICZ et al. (2000) discloses the interaction between the NS3 protease and a small peptide derived from NS4A. However,

BUTKIEWICZ et al. (2000) discloses exclusively *in vitro* assays for interactions between the small peptides derived from NS4A and the NS3 protease from HCV which may not be of physiological relevance.

There is a need in the art for polypeptides that contain the minimal aminoacid sequence that is able to bind specifically with a naturally-occurring HCV protein in physiological conditions in order to design new tools for therapeutic and detection purposes related to HCV.

SUMMARY OF THE INVENTION

This invention provides nucleic acids encoding polypeptides, which are termed SID® polypeptides, wherein these polypeptides are the final products of a double selection method involving a first step of selection of HCV-derived polynucleotides through a two-hybrid system and a second selection step involving an alignment between the different polynucleotides selected at the first step.

The invention also pertains to the SID® polypeptides encoded by the SID® nucleic acids.

Another object of the invention are recombinant vectors containing a SID® nucleic acid as defined above as well as host cells transformed with such vectors or nucleic acids.

A further object of the invention consists of two-hybrid methods which make use of these SID® nucleic acids as well as to methods for selecting molecules which inhibit the binding between a SID® polypeptide and a polypeptide that binds specifically thereto, as well as kits for performing these methods.

It is still a further object of the invention to provide for marker compounds which comprise a SID® polypeptide or which are encoded by a polynucleotide containing a SID® nucleic acid as defined above, as

well as to methods and kits which make use of these marker compounds.

This invention also relates to pharmaceutical compositions as well as to methods for preventing or curing a HCV viral infection in a human or an animal that use a SID® polypeptide or a SID® nucleic acid as disclosed herein.

Throughout this application, various publications, patents and published patent applications are cited. The disclosures of these publications, patents and published patent specifications, referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

BRIEF DESCRIPTION OF THE FIGURES.

Figure **1** consists of a general overview of HCV genome and its encoded polyprotein. The RNA coding strand is represented with a line for untranslated regions (NCR) and boxes for coding regions.

Positions and enzymes responsible for cleavage are indicated above. p7 is a secondary cleavage product of E2 (adapted from HOUGHTON, 1996).

Fig. **2** is a restriction map of the plasmid pAS2ΔΔ which may be used for producing a recombinant " Selected Interacting Domain (SID®) " polypeptide or a recombinant marker compound of the invention.

Fig. **3** is a restriction map of the plasmid pACTII which may be used for producing a recombinant " Selected Interacting Domain (SID®) ".

Fig. **4** is a restriction map of the plasmid pUT18 which may be used for producing a recombinant " Selected Interacting Domain (SID®) ".

DA.
Le August 2 2000

Fig. **5** is a restriction map of the plasmid pUT18C which may be used for producing a recombinant " Selected Interacting Domain (SID®) ".

Fig. **6** is a restriction map of the plasmid pT25 which may be used for producing a recombinant " Selected Interacting Domain (SID®) ".

Fig. **7** is a restriction map of the plasmid pKT25 which may be used for producing a recombinant " Selected Interacting Domain (SID®) ".

Fig.**8** is an illustration of the first step of selecting a SID® nucleic acid of the invention, wherein it is performed a selection of different sets of overlapping nucleic acids primarily selected through a two-hybrid method, in order to define pre-SID nucleic acids. Three fragments frg1, frg2 and frg3 of lengths l1, l2 and l3 respectively. Fragment l1 and l2 are clustered together if the length of intersection, l, is greater than 30% of l1 and l2. Fragment frg3 is grouped with fragments frg1 and frg2 if the length of intersection between frg1 and frg3, l', is greater than 30% of l1 and l3 and if the length of intersection between frg 2 and frg 3, l », is greater than 30% of l2 and l3.

Fig.**9** illustrates the selection of pre-SID® nucleic acid from a particular set of overlapping nucleic acids previously selected through a two-hybrid method. The pre-SID® is defined as the intersection of all the fragments (frg1-6) in a cluster.

Fig.**10** illustrates the selection of a SID® nucleic acid from the overlapping regions between two pre-SID nucleic acids. A SID® is defined if the length of overlap between two pre-SID®s, l, is greater than 30 bp. Further SID®s are defined by non-overlapping areas if their length (l') represents more than 30% of the length of one of the fragments which contributes to the corresponding pre-SID® (frg1-6).

Fig.**11** illustrates a further step of determining SID® nucleic acids after alignment of two overlapping SID nucleic acids identified according to figure 10. Fragments frg1' and frg2' contribute to both SID®1 and SID®2 (top panel). For each SID®, the number of fragments are counted and fragments are assigned to the SID® with the most fragments. The remaining fragments are re-analysed and a new SID® is

defined as the region of intersection of these fragments (bottom panel, SID®2' - fragment 3' and fragment 4'.

Fig.**12** illustrates a map of the vector pB5 which may be used in example 1.

Fig.**13** illustrates a map of the vector pP6 which may be used in example 1.

DETAILED DESCRIPTION OF THE INVENTION

The present invention firstly provides for nucleic acids encoding SID® polypeptides.

As generally used herein, a « bait » nucleic acid encodes a « bait » polypeptide. A polypeptide is termed a « bait » polypeptide when this polypeptide is used to select a formerly unknown « prey » nucleic acid encoding a « prey » polypeptide which binds selectively with said « bait » polypeptide. Indeed, a « prey » nucleic acid which has been selected for binding to a given bait polypeptide may be used in another selection method or in another round of the same selection method as a « bait » nucleic acid encoding a « bait » polypeptide for the purpose of selection of new prey nucleic acids, encoding prey polypeptides which bind selectively with said bait polypeptide, it being understood that the nucleic acid encoding said bait polypeptide was formerly selected from a population of prey nucleic acids.

SELECTED INTERACTING DOMAIN (SID®) POLYPEPTIDES AND METHODS FOR THEIR PREPARATION.

A selected interacting domain polypeptide that binds specifically to a polypeptide of interest is the result of a two-step screening procedure, wherein :

1) the first step consists of selecting and characterizing a collection of nucleic acids (prey nucleic acids) encoding polypeptides which bind specifically to a given bait polypeptide of interest;
and

2) the second step of the two-step procedure consists of determining the nucleic acid sequences which encode for SID® polypeptides after having generated sets of polynucleotides from the collection of nucleic acids selected at step 1).

As a result of the original two-step screening procedure disclosed hereunder, every nucleic acid finally selected encodes a « Selected Interacting Domain (SID®) " polypeptide which binds with a high specificity with the bait polypeptide of interest.

Step 1) Selecting prey nucleic acids

The first step of selecting a collection of nucleic acids encoding polypeptides which binds specifically to the bait polypeptide is carried out through a yeast two-hybrid system. The yeast two-hybrid system is designed to study protein-protein interactions *in vivo,* and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein.

According to the present invention, the first step of the procedure for selecting a Selected Interacting Domain (SID®) polynucleotide encoding a Selected Interacting Domain (SID®) polypeptide consists of the two-hybrid screening system described by Fromont-Racine et al. (1997) or the method described by FLAJOLET et al. (2000). The yeast two-hybrid system utilizes hybrid proteins to detect protein-protein interactions by means of direct activation of a reporter gene expression. In essence, the nucleic acids encoding the two putative protein partners, the bait polypeptide of interest and the prey polypeptide, are genetically fused to the DNA-binding domain of a transcription factor and to a transcriptional activation domain, respectively.

Construction of the prey HCV nucleic acids library.

Then, a genomic DNA library prepared from the genome of the pathogenic H77 strain of HCV (Yanagi et al., 1997), is constructed in the specially designed vector pP6 shown in figure 13 after ligation to suitable linkers, such that every genomic DNA insert is fused to a nucleotide sequence in the vector that encodes the transcription of domain of the Gal4 protein.

The polypeptides encoded by the nucleotide inserts of the genomic DNA library thus prepared are termed " prey " polypeptides in the context of the presently described selection method of prey nucleic acids.

Construction of the bait nucleic acids library

The DNA fragments obtained after nebulization of the HCV genomic DNA are also inserted in plasmid pB5 shown in figure 12 wherein these DNA inserts are fused to a polynucleotide encoding the DNA binding domain of the Gal4 protein and the recombinant vectors are used to transform *E. coli* cells. The transformed *E. coli* cells are grown and plasmid DNA is extracted and sequenced.

These plasmids which code in frame fusion proteins are used as bait plasmids. Bait plasmids thus consist of a collection of recombinant pB5 plasmids each containing inserted therein a DNA fragment from the H77 strain HCV genome encoding a polypeptide consisting of all or part of a HCV protein or alternatively a polypeptide consisting of all or part of two HCV proteins encoded by contiguous nucleic acid sequences of the HCV genome.

The selected HCV bait nucleic acids of the invention are referred to as the nucleotide sequences SEQ ID N°114 to 150.

The selected HCV bait polypeptides encoded by the nucleic sequences SEQ ID N°114 to 150 consist respectively of the aminoacid sequences SEQ ID N°77 to 113.

DA.
Le August 2 2000

Detectable marker genes are already present within the chromosomic yeast DNA and consist respectively of the His3 and LacZ genes, such as described by FROMONT-RACINE et al. (1997) or FLAJOLET et al. (2000).

Then, the collection of nucleic acid inserts contained in the collection of *E. Coli* cell clones containing the genomic DNA or HCV DNA library previously prepared are used to transform a first yeast strain, namely the Y187 *Saccharomyces cerevisiae* strain (phenotype:MATα, Gal4Δ, gal80Δ, ade2-101, His3, Leu2-3, -112 Trp1-901, Ura3-52, URA3::UASGAL1-LacZ Met).

The nucleic acid encoding the bait polypeptide of interest is inserted in the appropriate vector, said vector being used to transform a second yeast strain which may be the CG1945 (MATa Gal4-542 Gal180-538, Ade2-101, His3*200, Leu2-3, -112 Trp1-901 Ura3-52, Lys2-801, URA3::GAL4 17Mers (X3)-CyC1TATA-LacZ LYS2::GAL1 UAS-GAL1TATA-His3 CYH^R).

Then , the two yeast strains are mated to obtain a collection of mated cells.

The clones derived from the collection of mated cells above which are positive in an X-Gal overlay assay are those for which an interaction between the recombinant bait polypeptide and a polypeptide encoded by a nucleic acid insert originating from the HCV genomic library has occurred.

The clones derived from the collection of mated cells above may also be selected in the presence of histidine, and the positive clones are those for which an interaction between the recombinant bait polypeptide and a polypeptide encoded by a nucleic acid insert originating from the HCV genomic library has occurred.

In a further step, the prey nucleic acid inserts contained in the positively selected clones are amplified and sequenced.

Step 2:determination of the nucleic acid sequences encoding a Selected Interacting Domain (SID®) polypeptide which binds specifically to a bait polypeptide of interest.

This is the second step of the two step procedure defined above, which allows the precise selection of nucleic acids encoding the SID® nucleic acids of the present invention which are derived from the H77 strain HCV genome.

The SID® nucleic acid selection procedure, which is disclosed hereunder, has been specifically designed for the HCV genome which encodes for a single polyprotein and which thus comprises contiguous Open Reading Frames, said polyprotein being further processed to produce at least 10 mature structural and non-structural viral proteins.

Thus, the second selection step of the two-step procedure consists of a method for determining a polynucleotide encoding a Selected Interacting Domain (SID®) of a prey polypeptide of interest derived from HCV, which prey polypeptide interacts with a bait polypeptide, wherein said method comprises the steps of :

a) selecting, from the collection of prey polynucleotides obtained at the end of the first step of the two-step procedure described herein, all prey polynucleotides encoding a prey polypeptide capable of interacting with said bait polypeptide and containing a common nucleic acid fragment;

b) aligning the nucleotide sequences of the prey polynucleotides selected at step a) and gathering in one set or in a plurality of sets of sequences those nucleotide sequences which have sequences that overlap for more than 30% of their respective nucleic acid length, wherein each common overlapping nucleotide sequence in one set of sequences defines a sequence encoding a pre-SID® polypeptide (see Figures 8 and 9); and

c) aligning two sequences encoding two respective pre-SID® polypeptides (see Figure 10), and :

i) defining an overlapping nucleic acid sequence between the sequences encoding the two respective pre-SID® polypeptides as a

sequence encoding a SID® polypeptide, provided that the overlapping sequence is of at least 30 nucleotides in length;

ii) defining a non-overlapping nucleic acid sequence between the sequences encoding the two respective pre-SID® polypeptides as a sequence encoding a SID® polypeptide, provided that (1) said non-overlapping sequence has more than 30 nucleotides in length and (2) said non-overlapping sequence represents at least 30% in length of any one of the polynucleotides contained in the set of prey polynucleotides used for defining the sequence encoding each pre-SID® polypeptide.

This method may further comprise the steps of :

d) counting the number of overlapping prey polynucleotides contained in a first set of polynucleotides defining a sequence encoding a first SID® polypeptide;

e) counting the number of overlapping prey polynucleotides contained in a second set of polynucleotides defining a sequence encoding a second SID® polypeptide which overlaps with the sequence encoding the first SID® polypeptide;

f) determining which sequence among those encoding respectively the first SID® polypeptide and the second SID® polypeptide has been defined with the largest number of prey polynucleotides and selecting this set of prey sequences.

g) adding to the set of prey sequences selected at step f) those sequences that were contained in the set of prey sequences used for defining the sequence encoding the SID® polypeptide with the smallest number of prey sequences and which overlap with the sequence encoding the SID® polypeptide with the largest number of prey sequences.;

h) aligning the prey sequences added at step g) with the sequences already contained in the set of prey sequences which defined the sequence encoding the SID® polypeptide with the largest number of prey sequences;

i) defining an overlapping sequence between the whole sequences which were aligned in step h), wherein said overlapping sequence consists of a sequence encoding a SID® polypeptide. (See Figure 11).

The method for selecting a SID® nucleic acid encoding a SID® polypeptide is an object of the present invention, as well as any SID® nucleic acid or any SID® polypeptide which may be obtained by this selection method.

SID® nucleic acids of the invention

The SID® nucleic acids selected as described above starting from the genome of the H77 strain of HCV are the nucleic acid sequences of SEQ ID N°39 to 76 which encode the SID® polypeptides of SEQ ID N°1 to 38.

A first object of the invention consists of a nucleic acid which encodes a polypeptide selected from the group consisting of the aminoacid sequences SEQ ID N°1 to 38 or a variant thereof, and a sequence complementary thereto.

For the purposes of the present invention, a first polynucleotide is considered as being « complementary » to a second polynucleotide when each base of the first polynucleotide is paired with the complementary base of the second polynucleotide whose orientation is reversed. The complementary bases are A and T(or A and U), or C and G.

Preferably, any one of the nucleic acid or the polypeptides encompassed by the invention is under a purified or an isolated form.

The term "isolated" for the purposes of the present invention designates a biological material (nucleic acid or protein) which has been removed from its original environment (the environment in which it is naturally present).

For example, a polynucleotide present in the natural state in a plant or an animal is not isolated. The same polynucleotide separated from the adjacent nucleic acids in which it is naturally inserted in the genome of the plant or animal is considered as being "isolated".

Such a polynucleotide may be included in a vector and/or such a polynucleotide may be included in a composition and remains nevertheless in the isolated state because of the fact that the vector or the composition does not constitute its natural environment.

The term "purified" does not require the material to be present in a form exhibiting absolute purity, exclusive of the presence of other compounds. It is rather a relative definition.

A polynucleotide is in the "purified" state after purification of the starting material or of the natural material by at least one order of magnitude, preferably 2 or 3 and preferably 4 or 5 orders of magnitude.

"Isolated polypeptide" or "isolated protein" is a polypeptide or protein which is substantially free of those compounds that are normally associated therewith in its natural state (e.g., other proteins or polypeptides, nucleic acids, carbohydrates, lipids). "Isolated" is not meant to exclude artificial or synthetic mixtures with other compounds, or the presence of impurities which do not interfere with biological activity, and which may be present, for example, due to incomplete purification, addition of stabilisers, or compounding into a pharmaceutically acceptable preparation.

Variants of a selected interacting domain (SID®) polypeptide and nucleic acids encoding them.

As intended herein, a variant of a Selected Interacting Domain (SID®) polypeptide may be either a variant polypeptide of the Selected Interacting Domain (SID®) polypeptide or a polypeptide which is encoded by a nucleic acid variant of the polynucleotide encoding said Selected Interacting Domain (SID®) polypeptide.

Polynucleotides which encode a polypeptide variant of a Selected Interacting Domain (SID®) polypeptide, as the term is used herein, are polynucleotides that differ from the reference polynucleotide

encoding the parent SID® polypeptide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the reference polynucleotide may be generated by mutagenesis techniques, including those applied to polynucleotides, cells or organisms well known to one skilled in the art.

Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences which are at least 95% identical after optimal alignment to the reference polynucleotide of SEQ ID N°39 to 76 encoding the reference Selected Interacting Domain (SID®) polypeptide, preferably at least 96%, 97%, 98% and most preferably at least 99% identical to the reference polynucleotide. Similarly, a variant of a SID® polypeptide of the invention consists of a polypeptide having at least 95% aminoacid identity with a polypeptide selected from the aminoacid sequences SEQ ID N°1 to 38, and preferably at least 96%, 97%, 98% and most preferably at least 99% aminoacid identity with one of SEQ ID N°1 to 38.

Identity refers to sequence identity between two peptides or between two nucleic acid molecules. Identity between sequences can be determined by comparing a position in each of the sequences which may be aligned for purposes of comparison. When a position in the compared sequences is occupied by the same base or amino acid, then the sequences are identical at that position. A degree of identity between nucleic acid sequences is a function of the number of identical nucleotides at positions shared by these sequences. A degree of identity between amino acid sequences is a function of the number of identical aminoacids at positions shared by these sequences. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two

polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a " comparison window " to identify and compare local regions of sequence similarity. A " comparison window ", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for determining a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981), by the homology alignment algorithm of Needleman and Wunsch (1972), by the search for similarity method of Pearson and Lipman (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Solftware Package Release 7.0, Genetics Computer Group, 575, Science Dr. Madison, W1), or by inspection. The best alignment (i.e., resulting in the highest percentage of identity over the comparison window) generated by the various methods is selected. The term " sequence identity " means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term " percentage of sequence identity " is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g. A, T, C, G, U or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison

(i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

Most preferably, the percentage of nucleic acid or aminoacid identity between two nucleic acid or aminoacid sequences is calculated using the BLAST software (Version 2.06 of September 1998) with the default parameters.

Nucleotide changes present in a variant polynucleotide may be silent, which means that they do not alter the aminoacid encoded by the reference polynucleotide.

However, nucleotide changes may also result in aminoacid substitutions, additions, deletions, fusions and truncations in the Selected Interacting Domain (SD®) polypeptide encoded by the reference sequence.

The substitutions, deletions or additions may involve one or more nucleotides. Alterations may produce conservative or non-conservative aminoacid substitutions, deletions or additions.

Most preferably, the variant of a Selected Interacting Domain (SID®) polypeptide encoded by a variant polynucleotide possesses at least the same affinity of binding to its protein or polypeptide counterpart, against which it has been initially selected as described above.

The affinity of a given SID® polypeptide of the invention for a polypeptide into which it specifically binds is defined as the affinity constant Ka, wherein

$$Ka = \frac{[\text{SID®/polypeptide complex}]}{[\text{free SID®}]\,[\text{free polypeptide}]}$$

with [free SID®], [free polypeptide] and [SID®/polypeptide complex] consist of the concentrations at equilibrium respectively of the free SID® polypeptide, of the free polypeptide onto which the SID® polypeptide specifically binds and of the complex formed between the SID®

polypeptide and the polypeptide onto which said SID® polypeptide specifically binds.

Most preferably, the affinity of a SID® polypeptide of the invention or a variant thereof for its polypeptide counterpart (polypeptide partner) is assessed on a Biacore™ apparatus marketed by Amercham Pharmacia Biotech Company such as described by SZABO et al. (1995) and by Edwards and Leartherbarrow (1997).

As used herein, the expression « at least the same affinity» with reference to the affinity of binding between a SID® polypeptide of the invention to another polypeptide means that the Ka is identical or is of at least two-fold, preferably at least three-fold and most preferably at least five-fold greater than the Ka value of reference.

In another preferred embodiment, the variant of a Selected Interacting Domain (SID®) polypeptide which is encoded by a variant polynucleotide of the invention possesses a higher specificity of binding to its counterpart polypeptide or protein than the reference Selected Interacting Domain (SID®) polypeptide.

A variant of a Selected Interacting Domain (SID®) polypeptide according to the invention may be (1) one in which one or more, most preferably from one to three, of the aminoacid residues are substituted with a conserved or a non-conserved aminoacid residue and such substituted aminoacid residue may or may not be one encoded by the genetic code, or (2) one in which one or more of the aminoacid residues includes a substituent group.

In the case of an aminoacid substitution in the aminoacid sequence of a Selected Interacting Domain (SID®) polypeptide according to the invention, one or several-consecutive or non-consecutive - aminoacids are replaced by " equivalent " aminoacids. The expression " equivalent " aminoacid is used herein to designate any aminoacid that may be substituted for one of the aminoacids belonging to the native Selected Interacting Domain (SID®) polypeptide structure without decreasing the binding properties of the corresponding peptides to their counterpart polypeptide or protein, as regards the reference Selected Interacting Domain (SID®) polypeptide.

These equivalent aminoacids may be determined either by their structural homology with the initial aminoacids to be replaced, by the similarity of their net charge or of their hydrophobicity.

By an equivalent aminoacid according to the present invention is also meant the replacement of a residue in the L-form by a residue in the D-form or the replacement of a glutamic acid residue by a pyro-glutamic acid compound. The synthesis of peptides containing at least one residue in the D-form is, for example, described by KOCH (1977). A specific embodiment of a variant of a Selected Interacting Domain (SID®) polypeptide according to the invention includes, but is not limited to, a peptide molecule which is resistant to proteolysis, such as a peptide in which the -CONH- peptide bond is modified and replaced by a (-CH_2NH-) reduced bond, a (-NHCO-) retroinverso bond, a (-CH_2-O-) methylene-oxy bond, a (-CH_2-S-) thiomethylene bond, a (-CH_2CH_2-) carba bond, a (-CO-CH_2) hydroxyethylene bond, a (-N-N-) bond or also a -CH=CH bond.

As used herein, a variant of a SID® polypeptide of the invention also encompasses a polypeptide having an aminoacid sequence consisting of at least:

- 45 consecutive aminoacids of SEQ ID N°1;
- 30 consecutive aminoacidss of SEQ ID N°2;
- 65 consecutive aminoacids of SEQ ID N°3;
- 30 consecutive aminoacids of SEQ ID N°4;
- 130 consecutive aminoacids of SEQ ID N°5;
- 25 consecutive aminoacids of SEQ ID N°6;
- 23 consecutive aminoacids of SEQ ID N°7.
- 48 consecutive aminoacids of SEQ ID N°8;
- 36 consecutive aminoacids of SEQ ID N°9;
- 25 consecutive aminoacids of SEQ ID N°10;
- 24 consecutive aminoacids of SEQ ID N°11;
- 37 consecutive aminoacids of SEQ ID N°12;
- 25 consecutive aminoacids of SEQ ID N°13;
- 30 consecutive aminoacids of SEQ ID N°14;
- 27 consecutive aminoacids of SEQ ID N°15;
- 69 consecutive aminoacids of SEQ ID N°16;

- 130 consecutive aminoacids of SEQ ID N°17;
- 33 consecutive aminoacids of SEQ ID N°18;
- 25 consecutive aminoacids of SEQ ID N°19;
- 40 consecutive aminoacids of SEQ ID N°20;
- 78 consecutive aminoacids of SEQ ID N°21;
- 39 consecutive aminoacids of SEQ ID N°22;
- 57 consecutive aminoacids of SEQ ID N°23;
- 26 consecutive aminoacids of SEQ ID N°24;
- 68 consecutive aminoacids of SEQ ID N°25;
- 34 consecutive aminoacids of SEQ ID N°26;
- 42 consecutive aminoacids of SEQ ID N°27;
- 48 consecutive aminoacids of SEQ ID N°28.
- 102 consecutive aminoacids of SEQ ID N°29:
- 49 consecutive aminoacids of SEQ ID N°30:
- 92 consecutive aminoacids of SEQ ID N° 31;
- 49 consecutive aminoacids of SEQ ID N°30;
- 92 consecutive aminoacids of SEQ ID N°31;
- 71 consecutive aminoacids of SEQ ID N°32;
- 55 consecutive aminoacids of SEQ ID N°33;
- 69 consecutive aminoacids of SEQ ID N°34;
- 23 consecutive aminoacids of SEQ ID N°35;
- 33 consecutive aminoacids of SEQ ID N°36;
- 32 consecutive aminoacids of SEQ ID N°37;

and

- 22 consecutive aminoacids of SEQ ID N°38.

Without wishing to be bound by any particular theory, the inventors believe that polypeptides having an aminoacid length of about 10% lesser than the aminoacid length of anyone of the SID® polypeptides of SEQ ID N°1 to 39 of the invention have a high probability to retain the binding properties to a given (bait) polypeptide of the parent SID® polypeptide.

The invention also pertains to a nucleic acid encoding a SID® polypeptide which is selected from the group consisting of the sequences SEQ ID N°39 to 76, and a sequence complementary thereto.

The invention is also directed to a nucleic acid encoding a variant of SID® polypeptide selected from the group consisting of the sequences SEQ ID N°39 to 76, in reference to the definition of the SID® polypeptide variants above.

For example, a nucleic acid encoding a polypeptide having an aminoacid sequence consisting of at least 45 consecutive aminoacids of SEQ ID N°1 comprise at least 135 (45 x 3) consecutive nucleotides of the polynucleotide of SEQ ID N°39.

The same definition also apply for nucleic acids encoding variants of the SID® polypeptides of SEQ ID N°2 to 38, which are part of the invention.

The invention further relates to a nucleic acid encoding a polypeptide having an aminoacid sequence comprising from 1 to 3 substitutions, additions or deletions of one aminoacid as regards a polypeptide selected from the group consisting of the aminoacid sequences SEQ ID N°1 to 38 or a sequence complementary thereto.

Another object of the invention consists of a polypeptide selected from the group consisting of the aminoacid sequences SEQ ID N°39 to 76 or a variant thereof.

Are encompassed in the family of variants of a SID® polypeptide of the invention those polypeptides having an aminoacid sequence comprising from 1 to 3 substitutions, additions or deletions of one aminoacid as regards a polypeptide selected from the group consisting of the aminoacid sequences SEQ ID N°1 to 38.

The invention is also directed to an antibody directed against a a SID® polypeptide as defined above, or to a variant thereof.

The antibodies directed specifically against the Selected Interacting Domain (SID®) polypeptide or a variant thereof may be indifferently radioactively or non-radioactively labelled.

Monoclonal antibodies directed against a SID® polypeptide may be prepared from hybridomas according to the technique described

by Kohler and Milstein in 1975. Polyclonal antibodies may be prepared by immunization of a mammal, especially a mouse or a rabbit, with the SID® polypeptide that is combined with an adjuvant of immunity, and then by purifying the specific antibodies contained in the serum of the immunized animal on a affinity chromatography column on which has previously been immobilized the polypeptide that has been used as the antigen.

Antibodies directed against a SID® polypeptide may also be produced by the trioma technique and by the human B-cell hybridoma technique (Kozbor et al., 1983).

Antibodies directed to a SID® polypeptide include chimeric single chain Fv antibody fragments (US Patent N° US 4,946,778; Martineau et al., 1998), antibody fragments obtained through phage display libraries (Ridder et al., 1995) and humanized antibodies (Reinmann et al., 1997; Leger et al., 1997). Also, transgenic mice, or other organisms such as other mammals, may be used to express antibodies, including for example, humanized antibodies directed against a SID® polypeptide of the invention, or a variant thereof.

VECTORS OF THE INVENTION

The nucleic acids coding for a Selected Interacting Domain (SID®) polypeptide or a variant thereof, which are defined in the section above, can be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. Such transcription elements include a regulatory region and a promoter as defined previously. Thus, the nucleic acid encoding a marker compound of the invention is operably linked with a promoter in a expression vector, wherein said expression vector may include a replication origin.

The necessary transcriptional and translation of signals is most preferably provided by the recombinant expression vector.

Structure of the vectors encompassed by the invention

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acids of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic DNA sequences. Suitable vectors include derivatives of SV40 and known bacterial plasmids, e.g., *Escherichia coli* plasmids col El, pCR1, pBR322, pMal-C2, pET, pGEX (Smith *et al.*, 1988), pMB9 and their derivatives, plasmids such as RP4; phage DNAs, e.g., the numerous derivatives of phage l, e.g., NM989, and other phage DNA, e.g., M13 and filamentous single stranded phage DNA; yeast plasmids such as the 2m plasmid or derivatives thereof; vectors useful in eukaryotic cells, such as vectors useful in insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAs, such as plasmids that have been modified to employ phage DNA or other expression control sequences; and the like.

For example, in a baculovirus expression system, both non-fusion transfer vectors, such as but not limited to pVL941 (*Bam*H1 cloning site; Summers), pVL1393 (*Bam*H1, *Sma*I, *Xba*I, *Eco*R1, *Not*I, *Xma*III, *Bgl*II, and *Pst*I cloning site; Invitrogen), pVL1392 (*Bgl*II, *Pst*I, *Not*I, *Xma*III, *Eco*RI, *Xba*I, *Sma*I, and *Bam*H1 cloning site; Summers and Invitrogen), and pBlue*Bac*III (*Bam*H1, *Bgl*II, *Pst*I, *Nco*I, and *Hind*III cloning site, with blue/white recombinant screening possible; Invitrogen), and fusion transfer vectors, such as but not limited to pAc700 (*Bam*H1 and *Kpn*I cloning site, in which the *Bam*H1 recognition site begins with the initiation codon; Summers), pAc701 and pAc702 (same as pAc700, with different reading frames), pAc360 (*Bam*H1 cloning site 36 base pairs downstream of a polyhedrin initiation codon; Invitrogen(195)), and pBlueBacHisA, B, C (three different reading frames, with *Bam*H1, *Bgl*II, *Pst*I, *Nco*I, and *Hind*III cloning site, an N-terminal peptide for ProBond purification, and blue/white recombinant screening of plaques; Invitrogen (220) can be used.

Mammalian expression vectors contemplated for use in the invention include vectors with inducible promoters, such as the dihydrofolate reductase (DHFR) promoter, *e.g.*, any expression vector with a *DHFR* expression vector, or a *DHFR*/methotrexate co-

amplification vector, such as pED (*Pst*I, *Sal*I, *Sba*I, *Sma*I, and *Eco*RI cloning site, with the vector expressing both the cloned gene and *DHFR*; Kaufman,1991). Alternatively, a glutamine synthetase/methionine sulfoximine co-amplification vector, such as pEE14 (*Hind*III, *Xba*I, *Sma*I, *Sba*I, *Eco*RI, and *Bcl*I cloning site, in which the vector expresses glutamine synthase and the cloned gene; Celltech). In another embodiment, a vector that directs episomal expression under control of Epstein Barr Virus (EBV) can be used, such as pREP4 (*Bam*H1, *Sfi*I, *Xho*I, *Not*I, *Nhe*I, *Hind*III, *Nhe*I, *Pvu*II, and *Kpn*I cloning site, constitutive RSV-LTR promoter, hygromycin selectable marker; Invitrogen), pCEP4 (*Bam*H1, *Sfi*I, *Xho*I, *Not*I, *Nhe*I, *Hind*III, *Nhe*I, *Pvu*II, and *Kpn*I cloning site, constitutive hCMV immediate early gene, hygromycin selectable marker; Invitrogen), pMEP4 (*Kpn*I, *Pvu*I, *Nhe*I, *Hind*III, *Not*I, *Xho*I, *Sfi*I, *Bam*H1 cloning site, inducible methallothionein IIa gene promoter, hygromycin selectable marker: Invitrogen), pREP8 (*Bam*H1, *Xho*I, *Not*I, *Hind*III, *Nhe*I, and *Kpn*I cloning site, RSV-LTR promoter, histidinol selectable marker; Invitrogen), pREP9 (*Kpn*I, *Nhe*I, *Hind*III, *Not*I, *Xho*I, *Sfi*I, and BamHI cloning site, RSV-LTR promoter, G418 selectable marker; Invitrogen), and pEBVHis (RSV-LTR promoter, hygromycin selectable marker, N-terminal peptide purifiable via ProBond resin and cleaved by enterokinase; Invitrogen). Selectable mammalian expression vectors for use in the invention include pRc/CMV (*Hind*III, *Bst*XI, *Not*I, *Sba*I, and *Apa*I cloning site, G418 selection; Invitrogen), pRc/RSV (*Hind*III, *Spe*I, *Bst*XI, *Not*I, *Xba*I cloning site, G418 selection; Invitrogen), and others. Vaccinia virus mammalian expression vectors (*see*, Kaufman, 1991, *supra*) for use according to the invention include but are not limited to pSC11 (*Sma*I cloning site, TK- and b-gal selection), pMJ601 (*Sal*I, *Sma*I, *Afl*I, *Nar*I, *Bsp*MII, *Bam*HI, *Apa*I, *Nhe*I, *Sac*II, *Kpn*I, and *Hind*III cloning site; TK- and b-gal selection), and pTKgptF1S (*Eco*RI, *Pst*I, *Sal*I, *Acc*I, *Hind*II, *Sba*I, *Bam*HI, and Hpa cloning site, TK or XPRT selection).

Yeast expression systems can also be used according to the invention to express a Selected Interacting Domain (SID®) polypeptide or a variant thereof and also a marker compound as defined herein. For example, the non-fusion pYES2 vector (*Xba*I, *Sph*I, *Sho*I, *Not*I, *Gst*XI,

*Eco*RI, *Bst*XI, *Bam*H1, *Sac*I, *Kpn*1, and *Hind*III cloning sit; Invitrogen) or the fusion pYESHisA, B, C (*Xba*I, *Sph*I, *Sho*I, *Not*I, *Bst*XI, *Eco*RI, *Bam*H1, *Sac*I, *Kpn*I, and *Hind*III cloning site, N-terminal peptide purified with ProBond resin and cleaved with enterokinase; Invitrogen), to mention just two, can be employed according to the invention.

Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As previously explained, the expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (*e.g.*, lambda), and plasmid and cosmid DNA vectors, to name but a few.

Vectors are introduced into the desired host cells by methods known in the art, *e.g.*, transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (lysosome fusion), use of a gene gun, or a DNA vector transporter (see, *e.g.*, Wu et al., 1992; Wu and Wu, 1988; Canadian Patent Application No. 2,012,311, filed March 15, 1990).

A cell has been "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. A cell has been "transformed" by exogenous or heterologous DNA when the transfected DNA effects a phenotypic change.

For introducing a vector in a cell host, explicit reference is made to research carried out by the group of E. Wagner, relating to gene delivery by means of plasmid-polylysine complexes (Curiel et al., 1991; and Curiel et al., 1992). The plasmid-polylysine complex investigated upon exposition to certain cell lines showed at least some expression of the gene. Further, it was found that the expression efficiency increased considerably due to the binding of transferrin to the plasmid-polylysine complex. Transferrin gives rise to close cell-complex contact with cells comprising transferrin receptors; it binds the entire complex to the transferrin receptor of cells. Subsequently, at least part of the entire complex was found to be incorporated in the cells investigated.

Several different approaches have been developed for gene transfer. These include the use of viral based vectors (e.g., retroviruses, adenoviruses, and adeno-associated viruses) (Drumm, M. L. et al., Rosenfeld, M. A. et al., 1992; and Muzyczka, 1992), charge associating the DNA with an asialorosomucoid/poly L-lysine complex (Wilson, J. M. et al. 1992), charge associating the DNA with cationic liposomes (Brigham, K. L. et al., 1993) and the use of cationic liposomes in association with a poly-L-lysine antibody complex (Trubetskoy, V. S. et al., 1993).

Compositions comprising vectors of the invention.

Although non-viral based transfection systems have not exhibited the efficiency of viral vectors, they have received significant attention, in both in vitro and in vivo research, because of their theoretical safety when compared to viral vectors. Synthetic cationic molecules, have been reported which reportedly "coat" the nucleic acid through the interaction of the cationic sites on the transfection agent and the anionic sites on the nucleic acid. The positively charged coating reportedly interacts with the negatively charged cell membrane to facilitate the passage of the nucleic acid through the cell membrane by non-specific endocytosis. (Schofield, 1995) These compounds have, however, exhibited considerable sensitivity to natural serum inhibition, which has probably limited their efficiency in vivo as gene transfection agents. (Behr 1994)

A number of attempts have been made to improve the efficiency of lipid-like cationic transfection agents, some involving the use of polycationic molecules. For example, several transfection agents have been developed that contain the polycationic compound spermine covalently attached to a lipid carrier. (Behr, 1994), discloses a lipopolyamine and shows it to be more efficient at transfecting cells than single charge molecules (albeit still less efficient than viral vectors). The agent reported by Behr was, however, toxic, and caused cell death.

A few such lipid delivery systems for transporting DNA, proteins, and other chemical materials across membrane boundaries have been synthesized by research groups and business entities. Most of the synthesis schemes are relatively complex and generate lipid based delivery systems having only limited transfection abilities. A need exists in the field of gene therapy for cationic lipid species that have a high biopolymer transport efficiency. It has been known for some time that a very limited number of certain quaternary ammonium derivatized (cationic) liposomes spontaneously associate with DNA, fuse with cell membranes, and deliver the DNA into the cytoplasm (as noted above, these species have been termed "cytofectins"). LIPOFECTIN TM. represents a first generation of cationic liposome formulation development. LIPOFECTIN TM is composed of a 1:1 formulation of the quaternary ammonium containing compound DOTMA and dioleoylphosphatidylethanolamine sonicated into small unilamellar vesicles in water. Problems associated with LIPOCFECTIN TM include non-metabolizable ether bonds, inhibition of protein kinase C activity, and direct cytotoxicity. In response to these problems, a number of other related compounds have been developed. The monoammonium compounds of the subject invention improve upon the capabilities of existing cationic liposomes and serve as a very efficient delivery system for biologically active chemicals.

Most preferred vectors of the invention.

Most preferred recombinant vectors according to the invention include pASΔΔ(figure 2), pACTIIst (figure 3), pT18 (figure 4), pUT18C (figure 5), pT25 (figure 6), pKT25(figure 7), pB5 (Figure 12) and pP6 (Figure 13) containing inserted therein a nucleic acid encoding a Selected Interacting Domain (SID®) polypeptide or a variant thereof as defined above.

The present invention is also directed to a vector usable in a two-hybrid method which consists of the vector pP6 which is shown in figure 13. As disclosed in example 1, the vector pP6 has been successfully used for preparing a collection of recombinant plasmids

consisting of a genomic DNA library from the pathogenic strain H77 of the hepatitis C virus.

The invention also pertains to a vector usable in two-hybrid method which consists of the vector pB5. As disclosed in example 1, the vector pB5 has been successfully used in a yeast two hybrid method as a bait plasmid.

RECOMBINANT CELL HOSTS

In one embodiment, a Selected Interacting Domain (SID®) polypeptide of the invention or a variant thereof is recombinantly produced in a desired host cell which has been transfected or transformed with a nucleic acid encoding said Selected Interacting Domain (SID®) polypeptide or with a recombinant vector as defined above within which a nucleic acid encoding a Selected Interacting Domain (SID®) polypeptide of the invention is inserted.

Recombinant cell hosts are another aspect of the present invention.

Such cell hosts generally comprise at least one copy of a nucleic acid encoding a Selected Interacting Domain (SID®) polypeptide of the invention or a variant thereof

Preferred cells for expression purposes will be selected in function of the objective which is sought. For example, in the embodiment wherein the production of a Selected Interacting Domain (SID®) polypeptide according to the invention in large quantities is sought, the nature of the host cell used for its production is relatively indifferent, provided that large amounts of Selected Interacting Domain (SID®) polypeptides of the invention are produced and that optional further purification steps may be carried out easily.

However, in the embodiment wherein the Selected Interacting Domain (SID®) polypeptide is recombinantly produced within a host organism for the purpose of interfering with a specific protein-protein interaction, then the host organism is selected among the host

organisms which are suspected to produce naturally said polypeptide of interest.

Consequently, mammalian and typically human cells, as well as bacterial, yeast, fungal, insect, nematode and plant cells are cell hosts encompassed by the invention and which may be transfected either by a nucleic acid or a recombinant vector as defined above.

Examples of suitable recombinant host cells include VERO cells, HELA cells (e.g. ATCC N°CCL2), CHO cell-lines (e.g. ATCC N°CCL61) COS cells (e.g. COS-7 cells; COS cell referred to ATCC N°CRL1650), W138, BHK, HepG2, 3T3 (e.g. ATCC N°CRL6361), A549, PC12, K562 cells, 293 cells, Sf9 cells (e.g. ATCC N°CRL1711) and Cv1 cells (e.g. ATCC N°CCL70).

Other suitable host cells are usable according to the invention include prokaryotic host cells strains of *Escherichia coli* (e.g. strain DH5-α), of *Bacillus subtilis*, of *Salmonella typhimurium*, or strains of genera such as *Pseudomonas, Streptomyces* and *Staphylococcus*.

Further suitable host cells usable according to the invention include yeast cells such as those of *Saccharomyces*, typically *Saccharomyces cerevisiae*.

The invention also relates to a method for producing a SID® polypeptide as defined above, wherein said method comprises the steps of:

a) cultivating a cell host which has been transformed with a SID® nucleic acid of the invention or with a vector containing a SID® nucleic acid in an appropriate culture medium;

b) recovering the SID® recombinant polypeptide from the culture supernatant or from the cell lysate.

The SID® polypeptides or variant thereof thus recombinantly obtained may be purified, for example by high performance liquid chromatography, such as reverse phase and/or cationic exchange HPLC, as described by ROUGEOT et al. (1994). The reason to prefer this kind of peptide or protein purification is the lack of by-products found in the elution samples which renders the resultant purified protein more suitable for a therapeutic use.

TWO-HYBRID METHODS OF THE INVENTION

a) Yeast two-hybrid methods

The invention also pertains to a yeast two-hybrid method for selecting a recombinant cell clone containing a vector comprising a nucleic acid insert encoding a prey polypeptide which binds with a SID® polypeptide of SEQ ID N°1 to 38 or a variant thereof, wherein said method comprises the steps of :

a) mating at least one first recombinant yeast cell clone of a collection of recombinant yeast cell clones transformed with a plasmid containing the prey polynucleotide to be assayed with a second aploïd recombinant *Saccharomyces cerevisiae* cell clone transformed with a plasmid containing a bait polynucleotide encoding a SID® polypeptide of the invention or a variant thereof;

b) cultivating diploid cells obtained in step a) on a selective medium; and

c) selecting recombinant cell clones which grow on said selective medium.

The yeast two-hybrid method above may further comprise the step of :

d) characterizing the prey polynucleotide contained in each recombinant cell clone selected in step c).

Most preferably, such a yeast two-hybrid method may be performed by the one skilled in the art as it is disclosed in example 2 hereafter.

According to the yeast two-hybrid method above, a SID® polypeptide of the invention or a variant thereof is used as a bait polypeptide.

In a preferred embodiment of the yeast two-hybrid method described above, the prey polynucleotide is a DNA fragment from the genome of a pathogenic strain of the hepatitis C virus (HCV) ranging from about 150 to about 600 nucleotides in length and which is inserted in a vector which is contained in one recombinant clone of a collection of recombinant cell clones.

b) Bacterial two-hybrid method

A bacterial two-hybrid method of the invention may be performed by the one skilled in the art according to the teachings of KARIMOVA et al. (1998).

The first step of selecting a collection of nucleic acids encoding polypeptides which binds specifically to the bait polypeptide may also be carried out through a bacterial two-hybrid system.

According to such bacterial two-hybrid system, bacterial cell clones, preferably *Escherichia coli* cells, are transformed with a plasmid containing a bait polynucleotide encoding a bait polypeptide.

Then, plasmids containing a DNA insert are provided by rescuing the plasmids obtained from the collection of yeast clones containing the genomic DNA or cDNA library which are described in the previous section entitled " Yeast two-hybrid system ". For example, the plasmid rescue may be carried out according to the following steps:

(i) extracting plasmid DNA contained in the collection of yeast clones obtained as disclosed in the previous section, by using a conventional DNA extraction buffer and a phenol: chloroform: isoamyl alcohol (25:24:1) before centrifuging;

(ii) transferring a desired volume of the supernatant obtained at the end of step (i) to a sterile Eppendorf tube and add a precipitation buffer (ethanol/NH_4Ac) before centrifuging and resuspending the pellet after washing in ethanol;

(iii) transforming *Escherichia coli* cells (e.g. *Escherichia coli* cells of strain NC 1066) which have been rendered electrocompetent with a desired volume (e.g. 1 µl) of the yeast plasmid DNA extract obtained at step (ii) by electroporation;

(iv) collecting the transformed *Escherichia coli* cells.

Alternatively, a collection of *Escherichia coli* cell clones containing a collection of HCV genomic DNA inserts may be obtained by constructing the DNA library directly in the bacterial cell, such as disclosed in Flajolet et al. (2000).

Then, the bacterial recombinant cells which have been transformed both with a plasmid containing a bait polynucleotide encoding a bait polypeptide and a plasmid containing a prey polynucleotide encoding a prey polynucleotide is cultivated on a selective medium.

Then, recombinant cell clones capable of growing on said selective medium are selected and the DNA inserts of the plasmids containing therein are sequenced.

By bacterial two-hybrid system is generally intended a method that usually makes use of at least one reporter gene, the transcription of which is activated when a prey polypeptide and a bait polypeptide produced by the recombinant cell due to the triggering of the transcription of said at least one reporter gene when both the specific domain contained in one prey polypeptide and the complementary domain contained in the bait polypeptide are binding one to the other.

The invention further pertains to a bacterial two-hybrid method for identifying a recombinant cell clone containing a prey polynucleotide encoding a prey polypeptide which binds with a SID® polypeptide of SEQ ID N°1 to 38 or a variant thereof, wherein said method comprises the steps of :

a) transforming bacterial cell clones with a plasmid containing a SID® polynucleotide encoding a SID® polypeptide of the invention or a variant thereof;

b) rescuing prey plasmids containing prey polynucleotides wherein each prey polynucleotide is a DNA fragment from the genome of a desired organism and wherein each prey plasmid is contained in one recombinant yeast cell clone of a collection of recombinant yeast cell clones;

c) transforming the recombinant bacterial cell clones obtained in step a) with the plasmids rescued in step b);

d) cultivating bacterial recombinant cells obtained in step c) on a selective medium;

and

e) selecting recombinant cell clones which grow on said selective medium.

The bacterial two-hybrid system described above may further comprise the step of f) characterizing the prey polynucleotide contained in each recombinant cell clone selected at step e).

In one preferred embodiment of the yeast or bacterial two-hybrid methods described above, the prey polypeptide is a human polypeptide expressed by a mammal which is infected by the Hepatitis C virus, like human and monkeys, typically chimpanzees.

Generally, the yeast two-hybrid method or the bacterial two-hybrid method as disclosed herein may be performed with prey polypeptides of any origin, either of viral, fungal, bacterial or mammal origin, i.e. either of prokaryotic or eukaryotic origin.

In a second preferred embodiment of the two-hybrid methods above, the prey polypeptide is an HCV polypeptide.

Most preferably, the prey polypeptide is encoded by a strain of the hepatitis C virus which is pathogenic for human, such as strain H77.

SETS OF NUCLEIC ACIDS AND SETS OF POLYPEPTIDES OF THE INVENTION

In yet another aspect, the present invention relates to a set of two nucleic acids consisting of:

i) a first nucleic acid encoding a SID® polypeptide of SEQ ID N°1 to 39 of the invention or a variant thereof; and

ii) a second nucleic acid encoding a prey polypeptide which binds specifically with a SID® polypeptide defined in i).

In still a further aspect, the invention is also directed to a set of two polypeptides consisting of :

i) a first polypeptide consisting of a SID® polypeptide of SEQ ID N°1 to 39 of the invention or a variant thereof; and

ii) a second polypeptide which binds specifically with the first polypeptide.

The invention further relates to a complex formed between :

i) a first polypeptide consisting of a SID® polypeptide of SEQ ID N°1 to N°38 of the invention; and

ii) a second poplypeptide which binds specifically with the first polypeptide.

The invention also relates to a protein-protein interaction wherein the two interacting proteins consist of a set of two polypeptides as defined above.

In a preferred embodiment, the invention relates to the protein-protein interactions wherein the sets of two polypeptides consist of a SID® polypeptide of SEQ ID N°1 to 38 and an HCV polypeptide.

When several reiterations of the two-hybrid method are performed and thus common SID® polypeptide and prey polypeptides are selected, a map of all the interactions between these polypeptides may be designed, that take into account of the known and/or suspected biological function of each of the interacting polypeptides.

Table 1 illustrates protein-protein interaction between the SID® polypeptides of SEQ ID N°1 to 38 and polypeptides of SEQ ID N°77 to 113 which are encoded by the genome of strain H77 of the hepatitis C virus which is pathogenic for a mammal, like human or chimpanzee.

Thus, the data presented in table 1 disclose particular sets of nucleic acids as well as particular sets of polypeptides which are encompassed by the present invention.

For example, table 1 discloses that the nucleic acid of SEQ ID N°39 encodes the SID® polypeptide of SEQ ID N°1 which contains exclusively (100 %) an aminoacid sequence from the Core protein of HCV strain H77.

The nucleic acid of SEQ ID N°39 starts at the nucleotide in position 446 and ends at the nucleotide in position 600 of the HCV genome which is described by YANAGI et al. (1997).

Table 1 also discloses that the SID® polypeptide of SEQ ID N°1 is part of a set of polypeptides of the invention, wherein the second

polypeptide of said set of polypeptides consists of the polypeptide of SEQ ID N°77 which is encoded by the nucleic acid sequence of SEQ ID N°114, which nucleic acid sequence has 87% of its sequence which is derived from the region of the H77 strain HCV DNA encoding the Core protein.

Thus , a particular set of polypeptides according to the invention consists of:

i) the polypeptide of SEQ ID N°1; and

ii) the polypeptide of SEQ ID N°77.

The same reasoning apply for every set of polypeptides disclosed in table 1, which are expressly part of the present invention.

Similarly, a particular set of nucleic acids according to the invention consists of .

(i) the nucleic acid of SEQ ID N°39; and

(ii) the nucleic acid of SEQ ID N°114.

The same reasoning apply for every set of nucleic acids disclosed in table 1, which are expressly part of the present invention.

Thus, particular sets of two polypeptides of the invention are respectively SEQ ID N°77/SEQ ID N°1; SEQ ID N°78/SEQ ID N°2; SEQ ID N°78/SEQ ID N°3; SEQ ID N°79/SEQ ID N°4; SEQ ID N°80/SEQ ID N°5; SEQ ID N°81/SEQ ID N°6; SEQ ID N°82/SEQ ID N°7; SEQ ID N°83/SEQ ID N°8; SEQ ID N°84/SEQ ID N°9; SEQ ID N°85/SEQ ID N°10; SEQ ID N°86/SEQ ID N°11; SEQ ID N°87/SEQ ID N°12; SEQ ID N°88/SEQ ID N°13; SEQ ID N°89/SEQ ID N°14; SEQ ID N°90/SEQ ID N°15; SEQ ID N°91/SEQ ID N°16; SEQ ID N°92/SEQ ID N°17; SEQ ID N°93/SEQ ID N°18; SEQ ID N°94/SEQ ID N°19; SEQ ID N°95/SEQ ID N°20; SEQ ID N°96/SEQ ID N°21; SEQ ID N°97/SEQ ID N°22; SEQ ID N°98/SEQ ID N°23; SEQ ID N°99/SEQ ID N°24; SEQ ID N°100/SEQ ID N°25. SEQ ID N°101/SEQ ID N°26. SEQ ID N°102/SEQ ID N°27; SEQ ID N°103/SEQ ID N°28. SEQ ID N°104/SEQ ID N°29; SEQ ID N°105/SEQ ID N°30; SEQ ID N°106/SEQ ID N°31; SEQ ID N°107/SEQ ID N°32; SEQ ID N°108/SEQ ID N°33; SEQ ID N°109/SEQ ID N°34; SEQ ID N°110/SEQ ID N°35; SEQ ID N°111/SEQ ID N°36; SEQ ID N°112/SEQ ID N°37; and SEQ ID N°113/SEQ ID N°38.

Similarly, particular sets of two nucleic acids according to the invention are respectively: SEQ ID N°114/SEQ ID N°39; SEQ ID N°115/SEQ ID N°40; SEQ ID N°115/SEQ ID N°41; SEQ ID N°116/SEQ ID N°42; SEQ ID N°117/SEQ ID N°43; SEQ ID N°118/SEQ ID N°44; SEQ ID N°119/SEQ ID N°45; SEQ ID N°120/SEQ ID N°46; SEQ ID N°121/SEQ ID N°47; SEQ ID N°122/SEQ ID N°48; SEQ ID N°123/SEQ ID N°49; SEQ ID N°124/SEQ ID N°50; SEQ ID N°125/SEQ ID N°51; SEQ ID N°126/SEQ ID N°52; SEQ ID N°127/SEQ ID N°53; SEQ ID N°128/SEQ ID N°54; SEQ ID N°129/SEQ ID N°55; SEQ ID N°130/SEQ ID N°56; SEQ ID N°131/SEQ ID N°57; SEQ ID N°132/SEQ ID N°58; SEQ ID N°133/SEQ ID N°59; SEQ ID N°134/SEQ ID N°60; SEQ ID N°135/SEQ ID N°61; SEQ ID N°136/SEQ ID N°62; SEQ ID N°137/SEQ ID N°63; SEQ ID N°138/SEQ ID N°64; SEQ ID N°139/SEQ ID N°65; SEQ ID N°140/SEQ ID N°66; SEQ ID N°141/SEQ ID N°67; SEQ ID N°142/SEQ ID N°68; SEQ ID N°143/SEQ ID N°69; SEQ ID N°144/SEQ ID N°70. SEQ ID N°145/SEQ ID N°71; SEQ ID N°146/SEQ ID N°72. SEQ ID N°147/SEQ ID N°73; SEQ ID N°148/SEQ ID N°74; SEQ ID N°149/SEQ ID N°75 and SEQ ID N°150/SEQ ID N°76.

The protein-protein interactions disclosed in table 1 allows the design of a map of interactions between various polypeptides encoded by the genome of the H77 strain of HCV.

In such a Protein Interaction Map (PIM®) wherein each SID® polypeptide is linked to the bait polypeptide onto which it specifically binds, for example by an arrow.

Such a Protein Interaction Map (PIM®) may help the one skilled in the art to decipher a whole metabolical and/or physiological pathway that is functionally active within a pathogenic strain of HCV. Protein Interaction Map and computable version of PIM® are part of the present invention.

Therefore, in still another aspect, the present invention is directed to a computable readable medium (such as floppy disk, CD-ROM and all electronic or magnetic format which can be read by a computer) having stored thereon protein-protein interactions according to

the invention, preferably stored in a form of a Protein Interaction MAP, as shown, for example, in FROMONT-RACINE et al. (1997).

In a preferred embodiment, the invention comprises a computable readable medium as defined above, wherein the protein-protein interactions stored thereon are linked to annotated data base, for example through Internet.

In another preferred embodiment, the invention comprises a data bank containing the protein-protein interactions stored thereon, said data bank being available on a world-wide web site.

METHODS FOR SELECTING INHIBITORS OF PROTEIN-PROTEIN INTERACTIONS OF THE INVENTION

The transformed host cells as described above can also be used as models so as to study the interactions between a SID® polypeptide of the invention and its binding partner polypeptide, or between a SID® polypeptide of the invention and chemical or protein compounds which inhibit the binding between said SID® polypeptide and its binding partner polypeptide.

Example of a SID® polypeptide and its binding partner polypeptides are typically the sets of polypeptides of the invention which are described above.

In particular, the transformed host cells of the invention may be used for the selection of molecules which interact with a SID® polypeptide as described herein, as cofactor or as inhibitor, in particular a competitive inhibitor, or alternatively having an agonist or antagonist activity on the protein-protein interaction wherein said SID® polypeptide is involved. Preferably, the said transformed host cells will be used as a model allowing, in particular, the selection of products which make it possible to prevent and/or to treat pathologies induced by the hepatitis C virus.

Consequently, the invention also consists of a method for selecting a molecule which inhibits the protein-protein interaction of a set of two polypeptides as defined above, wherein said method comprises the steps of :

a) cultivating a recombinant host cell containing a reporter gene the expression of which is toxic for said recombinant host cell, said host cell being transformed with two vectors wherein:

i) the first vector contains a nucleic acid comprising a polynucleotide encoding a first hybrid polypeptide containing one of said two-polypeptides and a DNA binding domain;

ii) the second vector contains a nucleic acid comprising a polynucleotide encoding a second hybrid polypeptide containing the second of said two polypeptides and an activating domain capable of activating said toxic reporter gene when the first and the second hybrid polypeptides are interacting;

on a selective medium containing the molecule to be ested and allowing the growth of said recombinant host cell when the toxic reporter gene is not activated; and

b) selecting the molecule which inhibits the growth of the recombinant host cell defined in step a).

The invention is also directed to a method for selecting a molecule which inhibits the protein-protein interaction of a set of two polypeptides as defined above, wherein said method comprises the steps of :

a) cultivating a recombinant host cell containing a reporter gene the expression of which is toxic for said recombinant host cell, said host cell being transformed with two vectors wherein:

i) the first vector contains a nucleic acid comprising a polynucleotide encoding a first hybrid polypeptide containing one of said two polypeptides and the first domain of an enzyme;

ii) the second vector contains a nucleic acid comprising a polynucleotide encoding a second hybrid polypeptide containing the second of said two polypeptides and the second part of said enzyme capable of activating said toxic reporter gene when the first and the second hybrid polypeptides are interacting, said interaction recovering the catalytic activity of the enzyme;

on a selective medium containing the molecule to be tested and allowing the growth of said recombinant host cell when the toxic gene is not activated; and

b) selecting the molecule which inhibits the growth of the recombinant host cell defined in step a).

In a preferred embodiment, said toxic reporter gene that can be used for negative selection is URA3, CYH1 or CYH2 gene.

For example, a method for the screening of a molecule which inhibits the interaction between a SID® polypeptide of the invention with its binding protein counterpart may comprise the following steps:

- transform a permeabilized yeast cell with two vectors, respectively a first vector containing a SID® nucleic acid of the invention and a second vector containing a prey nucleic acid as defined in the present specification;
- plate on top agar the transformed permeabilized yeast cells above on square boxes;
- apply by spotting the candidate inhibitor molecules to test on top agar as soon as it is solidified;
- incubates, for example, overnight at 30°C, and
- select the inhibitor compounds that allow the growth of the transformed yeast cells.

The invention also provides for a kit for the screening of a molecule which inhibits the protein-protein interaction of a set of two polypeptides as defined above, wherein said kit comprises a recombinant host cell containing a reporter gene the expression of which is toxic for said recombinant host cell, said host cell being transformed with two vectors wherein:

i) the first vector contains a nucleic acid comprising a polynucleotide encoding a first hybrid polypeptide containing one of said two polypeptides and a DNA binding domain;

ii) the second vector contains a nucleic acid comprising a polynucleotide encoding a second hybrid polypeptide containing the second of said two polypeptides and an activating domain capable of activating said toxic reporter gene when the first and the second hybrid polypeptides are interacting.

Another object of the invention consists of a kit for the screening of a molecule which inhibits the protein-protein interaction of a set of two polypeptides as defined above, wherein said kit comprises a

recombinant host cell containing a reporter gene the expression of which is toxic for said recombinant host cell, said host cell being transformed with two plasmids wherein:

i) the first vector contains a nucleic acid comprising a polynucleotide encoding a first hybrid polypeptide containing one of said two polypeptides and the first domain of a protein;

ii) the second vector contains a nucleic acid comprising a polynucleotide encoding a second hybrid polypeptide containing the second of said two polypeptides and the second part of said protein capable of activating said toxic reporter gene when the first and the second hybrid polypeptides are interacting, said interaction recovering the activity of the protein. In the selection methods above, the transcription or activating domain and the DNA-binding domain may be derived from Gal4 and LexA respectively.

In the embodiment wherein the first domain is a first part of an enzyme and a complementary domain is a second part of the same enzyme, and wherein the proximity of the two parts of the enzyme restores the enzyme activity and activates a reporter gene, the two parts of the enzymes are most preferably the T25 and T18 polypeptides that form the catalytic domain of the *Bordetella pertussis adenylate cyclase*.

As an illustrative embodiment, the reporter gene is chosen among the group consisting of a nutritional gene or also a gene the expression of which is visualised by colorimetry such as His3, LacZ or both LacZ and His3.

MARKER COMPOUNDS OF THE INVENTION

The Selected Interacting Domain (SID®) polypeptides of SEQ ID N°1 to 38 of the invention and variants thereof defined in the present specification, and which bind specifically to a polypeptide of interest (e.g. a bait polypeptide), are useful as reagents for detecting, labelling, targeting or purifying specifically a polypeptide of interest, typically a polypeptide encoded by HCV, within a sample, since the SID® polypeptides possess properties that have never been reached using

conventional detection compounds, such as those of an antibody or an antibody fragment.

Firstly, the SID® polypeptides of the invention possess a high specificity of binding to the polypeptide of interest, since a SID® polypeptide consists of a portion of a larger polypeptide which binds in a highly specific manner to the polypeptide of interest in the natural environment within the eukaryotic cell infected by the Hepatitis C virus.

Secondly, the SID® polypeptide generally has a low molecular weight, generally from 3 kDa, and are thus easy to produce, on the one hand, and, on the other hand, can be easily introduced within a cell when the detection of the localisation or of the expression of the polypeptide of interest is sought. Moreover, the small size of a SID® polypeptide allows its passage through inner cell barriers such as the nucleus membrane, or the membranes surrounding the different cell organites.

Thus, a first object of the invention consists of a marker compound wherein said compound comprises :

a) a Selected Interacting Domain (SID®) polypeptide of the invention or a variant thereof that binds specifically to the polypeptide of interest; and

b) a detectable molecule bound thereto.

Such a marker compound is primarily useful for detecting, labelling or targeting a polypeptide of interest, for example a polypeptide of interest contained in a sample.

A detectable molecule according to the invention comprises, or alternatively consists of, any molecule which produces or can be induced to produce a signal. The detectable molecule can be a member of the signal producing system that includes the signal producing means .

The detectable molecule may be isotopic or non-isotopic. By way of example and not limitation, the detectable molecule can be part of a catalytic reaction system such as enzymes, enzyme fragments, enzyme substrates, enzyme inhibitors, co-enzymes, or catalysts. Part of a chromogen system such as fluorophores, dyes, chemiluminescers, luminescers, or sensitizers. A dispersible particle that can be non-magnetic or magnetic, a solid support, a liposome, a ligand, a receptor, a hapten radioactive isotope, and soforth.

It must be generally understood that the whole embodiments disclosed in the present specification involving a Selected Interacting Domain (SID®) polypeptide is straightfully applied also to any variant thereof.

Fluorescent detectable molecules

In one aspect of the marker compound according to the invention, the detectable molecule consists of a fluorescent molecule. Fluorescent moieties which are frequently used as labels are for example those described by Ichinose et al. (1991). Other fluorescent detectable molecules are fluorescing isothiocyanate (FITC) such as described by Shattil et al. (1987) or by Goding et al. (1986). The fluorescent detectable molecule may also comprise a phycoerythrin as taught by Goding et al. (1986), and Shattil et al. (1985). Other examples of fluorescent detectable molecules suitable for use as labels of a marker compound according to the invention are rhodamine isothiocyanate, dansyl chloride and XRITC.

Another fluorescent detectable molecule consists of the green fluorescent protein (GFP) of the jelly fish *Aequorea victoria*, and their numerous fluorescent protein derivatives.

The one skilled in the art may advantageously refer to the articles of CHALFIE et al. (1994) and of HEIM et al. (1994) which discloses the uses of GFP for the study of gene expression and protein localisation. The one skilled in the art may also refer to the article of Rizzuto et al. (1995) , which discusses the use of wild-type GFP as a tool for visualising subcellular organelles in cells, to the article of KAETHER and GERDES (1995), which reports the visualisation of protein transport along the secretary passway using wild-type GFP, the article of HU and CHENG (1995), which relates to the expression of GFP in plant cells and also to the article of Davis et al. (1995) which discloses the GFP expression in drosophilia embryos. For the use of several fluorescent variants of GFP, the one skilled in the art may refer to the article of Delagrave et al. (1995), as well as to the article of Heim et al. (1995). DNA encoding GFP is available commercially, for example from

Clontech in Palo Alto, California, USA. The one skilled in the art may use also humanized GFP genes such as those described in the US Patent N°6,020,192 and also the GFP protein disclosed in the US Patent N°5,941,084.

Another fluorescent protein that may be used in a marker compound according to the invention consists of the yellow fluorescent protein (YFP).

A further suitable luminescent protein consists of the luciferase protein.

Detectable molecules exhibiting a catalytic activity

In another embodiment of a detectable molecule included in a marker compound according to the invention, said detectable molecule is endowed with a catalytic activity and may thus consists of enzymes and catalytically active enzyme fragments. Some enzymatic labels are described in US Patent N°3,654,090. Such enzymes may be for example horse radish peroxydase (HRP), alkaline phosphatase or glutathione peroxydase which are well known from the one skilled in the art.

Enzymes, enzyme fragments, enzyme inhibitors, enzyme substrates, and other components of enzyme reaction systems can be used as detectable molecules. Where any of these components is used as a detectable molecule, a chemical reaction involving one of the components is part of the signal producing system.

Coupled catalysts can also involve an enzyme with a non-enzymatic catalyst. The enzyme can produce a reactant, which undergoes a reaction catalysed by the non-enzymatic catalyst or the non-enzymatic catalyst may produce a substrate (including co-enzymes) for the enzyme. The one skilled in the art may advantageously refer to the US Patent N°4,160 645 which disclose a white variety of non enzymatic catalysts, which may be employed, the appropriate portions of which are incorporated therein by reference.

The enzyme or co-enzyme employed provides the desired amplification by producing a product, which absorbs light, e.g., a tye, or

emits lights upon irradiation, e.g., a fluorescer. Alternatively, the catalytic reaction can lead to direct light emission, e.g., chemiluminescence. A large number of enzymes and co-enzymes for providing such products are described in the US Patents N°4,275,149, columns 19 to 23 and N°4,318,980, columns 10 to 14 which disclosures are incorporated herein by reference.

A number of enzyme combinations are set forth in US Patent N°4,275,149, columns 23 to 28 which disclosures are incorporated herein by reference.

When a single enzyme is used as the detectable molecule, or alternatively as comprised in the detectable molecule, such enzymes may find use are hydrolases, transferases, lyases, isomerases, ligases or synthetases and oxydoreductases.

Alternatively, luciferases may be used such as firefly luciferase and bacterial luciferase.

Primarily, the enzymes of choice, based on the I.U.B. classification are: (i) class 1. Oxydoreductases and (ii) class 3. Hydrolases. Most preferred oxydoreductases are (i) dehydrogenases of class 1.1, more particularly 1.1.1, 1.1.3. and 1.1.99 and (ii) peroxydases in class 1.11. of the hydrolases, particularly class 3.1., more particularly 3.1.3 and class 3.2, more particularly 3.2.1. are preferred.

Illustrative dehydrogenases include malate dehydrogenase, glucose-6-phosphate dehydrogenase and lactate dehydrogenase. Of the oxydases, glucose oxydases is exemplary. Of the peroxydases, horse radish peroxydase is illustrative. Of the hydrolases, alkaline phosphatases, β-glucosydase and lysozyme are illustrative.

Chemiluminescent detectable molecules

The detectable molecule comprised within the marker compound according to the invention may also consist in a chemiluminescent moiety. The chemiluminescent source involves a compound, which becomes electronically excited by a chemical reaction and may emit light which serves at as the detectable signal or donates energy to a fluorescent acceptor.

A diverse number of families of compounds have been found to provide chemiluminescent under a variety of conditions. When family of compounds is 2,3-dihydro-1,4-phtalazinedinone. The most utilised compound is luminol, which is the 5-amino analogue of the compound above. Other members of the family include the 5-amino-6,7,8-trimethoxy-and the dimethylamine-[ca]benzo analogue. These compounds can be made to luminance with alkaline hydrogen peroxyde or calcium hypochlorite and base.

Another family of compounds is the 2,4,5-triphenylimidazoles, with lophine as the common name for the parent product. Chemiluminescent analogues include para-dimethylamino- and para-methoxy-substituents. Chemiluminescents may also be obtained with geridinium esters, dioxetanes and oxalates, usually oxalyl active esters, e.g., p-nitrophenyl and a peroxide, e.g., hydrogen peroxide, under basic conditions. Alternatively, luciferins may be used in conjunction with luciferase or lucigenins.

Radioactive detectable molecules

In a further embodiment of a detectable molecule comprised in a marker compound according to the invention, said detectable molecule is radio-actively labelled such as with [^{3}H], [^{32}P], [and [^{125}I].

Colloïdal metal detectable molecules

In still a further embodiment, the detectable molecule comprised in a marker compound according to the invention may include a colloïdal metal particle. Colloïdal metals have been employed in immuno assays previously. Mostly, they consisted of either colloïdal iron or gold. The one skilled in the art may advantageously refer to the articles of Horisberger (1981) and Martin et al. (1990). In other case, the metals are chosen for their colour, i.e., their presence is determined by their colour or electron density under an electron microscope. Both the colour and electron density are directly proportional to the mass of the metal colloïd.

STRUCTURE OF THE MARKER COMPOUNDS OF THE INVENTION

In a first preferred embodiment of a marker compound of the invention, the detectable molecule is covalently bound to the Selected Interacting Domain (SID®) polypeptide of SEQ ID N°1 to SEQ ID N°38 or a variant thereof.

According to this specific embodiment, detectable molecules comprising fluorescent proteins such as GFP and YFP, enzymes or enzyme fragments such as alkaline phosphatase, glutathione peroxydase and horse radish peroxydase, chemiluminescent molecules, radioactive labels or colloidal metal particles will be preferred.

General methods that may be used by the one skilled in the art for covalently binding the detectable molecules to the Selected Interacting Domain (SID®) polypeptide are described in the numerous bibliographic references related to the preparation of the antibody conjugates used for carrying out immunoassays.

In a second preferred embodiment of a marker compound according to the invention, the detectable molecule is non-covalently bound to the Selected Interacting Domain (SID®) polypeptide or a variant thereof.

In a first preferred aspect of this second preferred embodiment, the detectable molecule consists of an antibody directed specifically against the Selected Interacting Domain (SID®) polypeptide or a variant thereof.

The antibodies directed specifically against the Selected Interacting Domain (SID®) polypeptide or a variant thereof may be indifferently radioactivity or non radioactivity labelled.

NUCLEIC ACIDS ENCODING A MARKER COMPOUND OF THE INVENTION.

The present invention also relates to a nucleic acid encoding a marker compound as defined above.

Most preferred nucleic acids encompassed by the invention include polynucleotides that encode a marker compound wherein the Selected Interacting Domain (SID®) polypeptide of SEQ ID N°1 to 38 or a variant thereof is covalently bound to the detectable molecule and wherein the detectable molecule consists itself of a polypeptide.

Most preferred nucleic acids are those of SEQ ID N°39 to 76.

In a first preferred embodiment of a nucleic acid according to the invention, said nucleic acid encodes for a Selected Interacting Domain (SID®) polypeptide which is fused to a fluorescent protein, such as GFP and YFP.

In a second preferred embodiment of a nucleic acid according to the invention, said nucleic acid encodes for a Selected Interacting Domain (SID®) polypeptide which is fused to a polypeptide endowed with a catalytic activity, such as an enzyme or an enzymatically active enzyme fragment, like alkaline phosphatase, glutathione peroxydase and horse radish peroxydase.

In a preferred embodiment, a nucleic acid encoding a marker compound of the invention comprises a DNA coding sequence which is transcribed and translated into said marker compound in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon and a translation stop codon. A coding sequence can include, but is not limited to:

- prokaryotic sequences, for example when the Selected Interacting Domain (SID®) nucleic acid and the nucleic acid fused thereto which encodes the detectable molecule are of prokaryotic origin;

- prokaryotic and eukaryotic sequences, for example the nucleic acid encoding the detectable molecule originates from an eukaryotic host organism.

If the coding sequence is intended for expression in an eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

In a most preferred embodiment of a nucleic acid sequence according to the invention, said nucleic acid sequence include a

regulatory region which is functional in the host organism within which the expression of said nucleic acid sequence is sought, wherein said regulatory region comprises a promoter sequence.

"Regulatory region" means a nucleic acid sequence which regulates the expression of a nucleic acid. A regulatory region may include sequences which are naturally responsible for expressing a particular nucleic acid (a homologous region), or may include sequences of a different origin (responsible for expressing different proteins or even synthetic proteins). In particular, the sequences can be sequences of eukaryotic or viral genes or derived sequences which stimulate or repress transcription of a gene in a specific or non-specific manner and in an inducible or non-inducible manner. Regulatory regions include origins of replication, RNA splice sites, enhancers, transcriptional termination sequences, signal sequences which direct the polypeptide into the secretary pathways of the target cell, and promoters.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced and translated into the protein encoded by the coding sequence.

Most preferred vectors for the expression of a marker compound of the invention.

Most preferred recombinant vectors for expressing a marker compound of the invention include pASΔΔ (figure 2), pACTIIst (figure 3), pT18 (figure 4), pUT18C (figure 5), pT25 (figure 6), pKT25 (figure 7), pB5 (Figure 12) and pP6 (Figure 13) containing inserted therein a nucleic acid encoding a Selected Interacting Domain (SID®) polypeptide as defined above or a variant thereof.

The invention also pertains to recombinant host cells transformed with a vector expressing a marker compound as defined above, more particularly a vector comprising inserted therein a nucleic acid encoding said marker compound, which is operably linked to suitable regulation signals which are functional in the host cell wherein its expression is sought.

Preferred cells for expression purposes will be selected in function of the objective which is sought. For example, in the embodiment wherein the production of a marker compound according to the invention in large quantities is sought, the nature of the cell host used for its production is relatively indifferent, provided that large amounts of Selected Interacting Domain (SID®) polypeptides or marker compounds of the invention are produced and that optional further purification steps may be carried out easily.

However, in the embodiment wherein the marker compound is recombinantly produced within a host organism for the purpose of qualitative or quantitative analysis of the polypeptide of interest onto which said marker compound specifically binds, then the host organism is selected among the host organisms which are suspected to produce naturally said polypeptide of interest.

Consequently, mammalian and human cells, as well as bacterial, yeast, fungal, insect, nematode and plant cells are cell host encompassed by the invention and which may be transfected either by a nucleic acid or a recombinant vector as defined above.

DETECTION METHODS OF THE INVENTION

The present invention further relates to the use of a Selected Interacting (SID®) polypeptide of SEQ ID N°1 to 38 or a variant thereof as well as a nucleic acid encoding it for detection purposes such as nucleic acids of SEQ ID N°39 to 76. It is herein reminded that a Selected Interacting Domain (SID®) polypeptide is determined according to the ability of such a (SID®) polypeptide to bind in a highly specific manner to a given (e.g. bait) polypeptide of interest, since the aminoacid sequence of a SID® polypeptide is encoded by a nucleic acid, the nucleotide sequence of which consists of the polynucleotide sequence which is common to a collection of nucleic acid sequences encoding prey polypeptides that have been selected for their specific binding properties to a (bait) polypeptide of interest, such as explained above in the section entitled " SELECTED INTERACTING DOMAIN (SID®) POLYPEPTIDES ".

The specific properties of a Selected Interacting Domain (SID®) polypeptide for binding to a given polypeptide of interest, either a viral, yeast, fungal, bacterial, insect, plant or mammal polypeptide, including a polypeptide of human origin, allow its use as a specific ligand for said polypeptide of interest of which the detection is sought.

Therefore, the use of a Selected Interacting Domain (SID®) in any detection method known in the art and which makes use of the ability of a detection ligand to bind specifically to a molecule of interest, most preferably a polypeptide of interest, fall under the scope of the present invention.

Detection methods that make use of the recognition of a molecule of interest, most preferably a polypeptide of interest, by a detection ligand are well known in the art and are primarily illustrated by the abundant literature that relate to immunoassays, which is incorporated herein by reference in its entirety.

The one skilled in the art may particularly refer to the book of Maggio (1980) (Heterogeneous assays), the US Patent N°3,817,837 (homogeneous Immunoassays), US Patent N° 3,993,345 (Immunofluorescense methods), US Patent N°4,233,402 (enzyme channelling techniques), US Patent N°3,817,837 (Enzyme multiplied immunoassay technique), US Patent N°4,366,241 and European Patent

Application N°EP-A 0 143 574 (Migration type assays), US Patent N°5,202,006, US Patent N°5,120,413 and US Patent N°5,145,567 (Immunofixation electrophoresis, mmunoelectrophoresis), the article of Aguzzi et al. (1977), the article of White et al. (1986), the article of Merlini et al. (1983), the US Patent n°5,228,960 (Immunosubstraction electrophoresis), the articles of Chen et al. (1991), Nielsen et al. (1991) and the US Patent n° 5,120,413 (Capillary electrophoresis).

Acellular detection method of the invention.

A first detection method of the invention consists of a method for detecting a polypeptide of interest within a sample, wherein said method comprises the steps of:

a) contacting a marker compound or a plurality of marker compounds according to the invention with the sample which is suspected to contain the polypeptide of interest the detection of which is sought;

b) detecting the complexes formed between said marker compound or said plurality of marker compounds and said polypeptide of interest.

The sample which is assayed for the presence of the polypeptide of interest the detection of which is sought may be of any nature , including every sample that may be used for carrying out an immunoassay.

In a first aspect, the sample may be any biological fluid, such as blood or blood separation products (e.g. serum, plasma, buffy coat), urine, saliva, tears.

In a second aspect, the sample may be any isolated biological tissue sample, including tissue sections previously fixed for purposes of histological studies.

In a third aspect, the sample may be a culture supernatant of a cell culture and a cell lysate of cultured cells.

In a first preferred embodiment of the first detection method of the invention described above, the detection step b) consists of the measure of the fluorescence signal intrinsically emitted by the detectable

molecule. It may for exampole be taken the advantage of SID® polypeptides or variants thereof having in their aminoacid sequence one or several tryptophan aminoacid residues.

In a second preferred embodiment of the first detection method of the invention detailed above, the detection step b) consists of submitting the detectable molecule to a source of energy at the excitation wavelength of said detectable molecule, and measuring the light emitted at the emission wavelength of said detectable molecule.

An illustrative example of this second embodiment above is when the marker compound used consists of a Selected Interacting Domain (SID®) which is bound to a fluorescent molecule, such as the fluorescent proteins GFP or YFP.

For example, in the embodiment wherein the detectable molecule of the marker compound of the invention which is used according to the first detection method above comprises, or alternatively consists of, a GFP protein, the detection step c) includes illuminating the sample tested at an emission wavelength substantially equal to 490 nm, and measuring the light emitted by the marker compound which is bound to the polypeptide of interest within the sample at an emission wavelength substantially equal to 510 nm.

Preferably, the marker compounds which are not bound to the polypeptide of interest the detection of which is sought within the sample are removed before carrying out the detection step.

In a third preferred embodiment, the detection step c) of the first detection method of the invention consists of measuring the catalytic activity of the detectable molecule. In this specific embodiment, the marker compound used in the detection method comprises a detectable molecule which comprises, or alternatively which consists of, an enzyme or a catalytically active enzyme fragment, such as already detailed in the section entitled " Marker compounds of the invention ".

In a fourth preferred embodiment, the detection step b) consists of measuring the radioactivity emitted by the detectable molecule.

The present invention further relates to a kit for detecting a polypeptide of interest within a sample, wherein said kit comprises a marker compound according to the invention.

Optionally, said detection kit further comprises the reagents necessary for carrying out the detection step b), such as a suitable substrate for the particular enzyme or a catalytically active enzyme fragment used, as well as suitable buffer solutions, which may be identical to those conventionally used for performing immunoassays.

Cellular detection assay using a recombinantly produced marker compound of the invention.

As already described above, any marker compound according to the invention may be produced according to genetic engineering techniques. Particularly, nucleic acid encoding a particular marker compound which binds specifically to a polypeptide of interest the detection of which is sought may be inserted in a vector, wherein said vector may be used to transfect or transform a host organism, either a prokaryotic or an eukaryotic cell host such as defined above.

In this specific embodiment, the production of a recombinant marker compound of the invention is allowed within such a transfected or transformed host cell. Once the host cell of interest is transfected or transformed with such a recombinant vector and once the recombinant marker compound is produced within the cell host of interest, then the Selected Interacting Domain (SID®) polypeptide portion of said marker compound will be able to bind specifically to its specific target polypeptide within the cell host. In this situation, the recombinantly produced marker compound of the invention will predominantly be localised at cell sites wherein the targeted polypeptide of interest is present.

This is the purpose of the second detection method of the invention which is detailed below.

A further object of the invention consists of a method for detecting a polypeptide of interest within a prokaryotic or an eukaryotic cell host, wherein said method comprises the steps of :

a) providing a cell host to be assayed;

b) transfecting said cell host with a nucleic acid encoding a marker compound of the invention, or with a recombinant vector encoding a marker compound of the invention;

c) detecting the complexes formed between the marker compound expressed by the transfected cell host and the polypeptide of interest.

Because the Selected Interacting Domain (SID®) polypeptide which is part of a marker compound of the invention specifically binds to a polypeptide which is suspected to be naturally produced by the targeted cell host, the second detection method of the invention defined above allows a qualitative as well as a quantitative detection of this targeted polypeptide which is suspected to be naturally produced by the transfected target cell host under assay.

For example, in the embodiment within which the procedure for selecting the Selected Interacting Domain (SID®) polypeptide which is part of a marker compound of the invention includes a first step wherein a collection of clones containing nucleic acid inserts derived from a H77 strain HCV genomic DNA library is prepared, the transfection of a mammalian cell, preferably a human cell, with a vector encoding such a marker compound of the invention will allow to detect the expression of a human polypeptide naturally expressed within said mammalian host cell and which naturally interacts with the HCV viral protein from which is derived the Selected Interacting Domain (SID®) polypeptide.

The second detection method of the invention defined above firstly allows the qualitative detection of the targeted polypeptide of interest which binds specifically with the recombinantly produced marker compound of the invention, and thus permits to know in which environmental conditions or at which differentiation stage the targeted polypeptide of interest is naturally produced within the cell host transfected with a vector expressing a marker compound of the invention.

Secondly, this second detection method of the invention allows the localisation of the targeted polypeptide of interest within the interior

of the cell, including localisation in the plasma membrane, cytosol, nucleus and any organelle such as ribosomes, Golgi apparatus, lysosomes, phagosomes, endoplasmic reticulum and chloroplasts.

The localisation of a targeted polypeptide of interest which is expressed within the cell host under assay according to the second detection method of the invention may be carried out by any means well known in the art, including using a confocal microscope.

Thirdly, the second detection method of the invention allows also a quantitative analysis of the expression of the targeted polypeptide of interest within the cell host under assay, since the level of the detection signal produced by the detectable molecule which is part of the marker compound will be proportional to the number of complexes formed between the cell host under assay between the targeted polypeptide of interest and the recombinantly produced marker compound of the invention.

Essentially, the one skilled in the art may refer to the section entitled " Acellular detection method of the invention " above to find the teachings necessary for performing the detection step c) of the second detection method described herein.

In a first embodiment of said second detection method of the invention, the detection step c) consists of the measure of the fluorescence signal intrinsically emitted by the detectable molecule comprised in the recombinantly expressed marker compound of the invention.

In a second preferred embodiment of the second detection method above, the detection step c) consists of submitting the detectable molecule to a source of energy at the excitation wavelength of said detectable molecule and measuring the light emitted at the emission wavelength of said detectable molecule.

In still a further embodiment of the second detection method of the invention, the detection step c) consists of measuring the catalytic activity of the detectable molecule.

In another embodiment, the detection step c) consists of measuring the radioactivity emitted by the detectable molecule.

In yet a further embodiment of the second detection method of the invention, the detection step c) allows the location of the complexes formed between the recombinantly produced marker compound and the targeted polypeptide of interest within the transfected cell host.

A further object of the invention consists of a kit for detecting a polypeptide of interest within a prokaryotic or an eukaryotic cell host, wherein said kit comprises a nucleic acid encoding a marker compound as defined herein, or a recombinant vector containing inserted therein a nucleic acid encoding a marker compound of the invention.

Optionally, the detection kit above may further comprise the reagents necessary to carry out the detection step c).

Cellular detection method of the invention using a marker compound which is introduced within a cell host.

There is a third detection method according to the invention wherein the marker compound comprising a Selected Interacting Domain (SID®) polypeptide OF SEQ ID N°1 to 38 or a variant thereof is previously produced by any means and subsequently introduced into a target cell host for the purpose of detecting a targeted polypeptide of interest which binds specifically with said Selected Interacting Domain (SID®) polypeptide.

Thus, the invention further relates to a method for detecting a polypeptide of interest within a prokaryotic or an eukaryotic cell host, wherein said method comprises the step of :

a) providing a cell host to be assayed;

b) introducing a marker compound as defined herein within said cell host;

and

c) detecting the complexes formed between the marker compound and the polypeptide of interest within the cell host.

Taking into account the low molecular weight of the Selected Interacting Domain (SID®) polypeptide selected from SEQ ID N°1 to 38 which is part of a marker compound of the invention, when compared with conventional specific detection molecules such as antibodies or antibody fragments, it results that the introduction of a marker compound of the invention into the interior of a target cell host will be much more easier to perform, as compared with the introduction within a cell host of a conventional marker like a labelled antibody or a labelled antibody fragment.

According to the third detection method of the invention defined above, step b) of introducing the marker compound within the target cell host may be performed by any technique well known in the art, including electroporation, and the use of molecules that will facilitate the passage of the marker compound of the invention through the cell membranes, and typically the plasma membrane.

Such molecules that facilitate the passage of a marker compound of the invention through cell membranes include, but are not limited to, penetratin, like penetratin 1.RTM (Encor, Gaithersburg, Md), Antenna Pediae protein, cationic lipids and cationic polyacrylates.

Permeation enhancers which may be employed include bile salts such as sodium glycocholate and other molecules such as β-cyclodextrin. Bile salts are known to increase the absorption of macromolecules across membranes (Pontiroli et al., 1987).

As already detailed for the second detection method of the invention described in the previous section, the third detection method of the invention allows also the localisation of the targeted polypeptide of interest which is expressed by the cell host under assay, as well as the qualitative and quantitative analysis of the expression of said target polypeptide of interest.

The detection step c) according to the third detection method of the invention described above may be carried out in the same way than the detection step c) of anyone of the first detection method and the second detection method detailed in the previous sections herein.

In a first embodiment of the third detection method above, the detection step c) consists of the measure of the fluorescence signal intrinsically emitted by the detectable molecule.

In a second embodiment, the detection step c) consists of submitting the detectable molecule to a source of energy at the excitation wavelength of said detectable molecule and measuring the light emitted at the emission wavelength of said detectable molecule.

In a third embodiment, the detection step c) consists of measuring the catalytic activity of the detectable molecule.

In a fourth embodiment, the detection step c) consists of measuring the radioactivity emitted by the detectable molecule.

In a fifth embodiment of the third detection method of the invention, the detection step c) allows the location of the complexes formed between the marker compound and the polypeptide of interest within the target cell host under assay.

A further object of the invention consists of a kit for detecting a polypeptide of interest within a prokaryotic or an eukaryotic cell host, wherein said kit comprises a marker compound as defined herein.

The detection kit above may further comprise the reagents necessary to carry out the detection step c).

The detection kit above may also further comprise the reagents necessary to facilitate the introduction of the marker compound within the target cell host under assay.

SOLID PHASE DETECTION METHOD USING A SELECTED INTERACTING DOMAIN (SID®) POLYPEPTIDE.

In a further aspect of the invention, the use of a Selected Interacting Domain (SID®) polypeptide of SEQ ID N°1 to 38or a variant thereof for detection purpose include a step wherein said Selected Interacting Domain (SID®) polypeptide is immobilised on a suitable substrate before bringing a sample to be assayed in contact with the substrate onto which said Selected Interacting Domain (SID®) polypeptide has been previously immobilised.

A subsequent step will consist in detecting the complexes formed between the Selected Interacting Domain (SID®) polypeptide immobilised on the substrate and the targeted polypeptide of interest the presence of which is suspected in the sample assayed.

Thus, the invention also pertains to a fourth detection method which consists of a method for detecting a polypeptide or a plurality of polypeptides of interest within a sample, wherein said method comprises the steps of :

a) providing a substrate onto which a Selected Interacting Domain (SID®) polypeptide or a plurality of Selected Interacting Domain (SID®) polypeptides is (are) immobilised;

b) bringing into contact the substrate defined in a) with the sample to be assayed;

c) detecting the complexes formed between the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected

Interacting Domain (SID®) polypeptides and the target polypeptide or the plurality of target polypeptides contained in the sample.

Substrates, supports or surfaces for immobilising protein molecules are well known in the art, and a lot of them have been described for performing solid phase immunoassays.

Preferably, a plurality of Selected Interacting Domain (SID®) polypeptides of different aminoacid sequences choosen among the sequences SEQ ID N°1 to 38 are immobilised on the substrate used according to the fourth detection method of the invention.

For example, a complete collection of Selected Interacting Domain (SID®) polypeptides which have been determined according to the methods described in the section entitled " Selected Interacting Domain (SID®) polypeptides " above, using nucleic acids derived from the H77 strain HCV genomic DNA as starting material, may be used for being immobilised on a suitable substrate.

According to this embodiment, the collection of Selected Interacting Domain (SID®) polypeptides of SEQ ID N°1 to 38 are immobilised on the substrate in another manner, thus forming an ordered area of SID® polypeptides immobilised at known locations of the surface of said substrate.

The substrate, support or surface may be a porous or a non-porous water insoluble material. The support can be hydrophilic or capable of being rendered hydrophilic and includes inorganic powders such as silica, magnesium sulphate, and alumina; natural polymeric materials, particularly cellulosic materials and materials derived from cellulose, such as fiber containing papers; synthetic or modified naturally occurring polymers, such as nitro-cellulose, cellulose acetate, poly(vinyl chloride), polyacrylamide , cross-linked dextran, agarose, polyacrylate, polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polymethacrylate, poly(ethylene terephtalate), nylon, poly(vinyl butyrate), said materials being used by themselves or in conjunction with other materials; glass available as Bioglass, ceramic metals and the like.

An ordered area onto which a plurality of Selected Interacting Domain (SID®) polypeptides are immobilised may be manufactured

according to the techniques disclosed in the US Patent N°5,143,854 or the PCT Application n°WO 92/10092, incorporated herein by reference for all purposes. The combination of photolithographic and fabrication techniques may, for example, enable each Selected Interacting Domain (SID®) polypeptide to occupy a very small area (" site ") on the support. In some embodiments, the site may be as small as few microns or even a single Selected Interacting Domain (SID®) polypeptide.

In a first embodiment of the fourth detection method detailed above, the plurality of Selected Interacting Domain (SID®) polypeptides are immobilized on the substrate in an order manner.

In a second embodiment of Selected Interacting Domain (SID®), the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected Interacting Domain (SID®) polypeptides are covalently bound to the substrate.

In a third embodiment of said method, the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected Interacting Domain (SID®) polypeptides are non-covalently bound to the substrate. According to this specific embodiment, the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected Interacting Domain (SID®) polypeptides are covalently bound to a first ligand molecule and the substrate is coated with a second ligand molecule, wherein said second ligand molecule specifically binds to the first ligand molecule. According to such a specific embodiment, the first ligand may be biotin in which case the second ligand is most preferably streptavidin.

In still a further embodiment of the fourth detection method according to the invention, the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected Interacting Domain (SID®) polypeptides are covalently linked to a spacer, which spacer is itself also covalently bound to the substrate in order to immobilise the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected Interacting Domain (SID®) polypeptides onto said substrate. Such a spacer may be a peptide polymer such as a poly-alanine or a poly-lysine peptide of 10 to 15 amino acids in length.

In still a further embodiment of the fourth detection method above, the detection step c) consists of detecting changes in the optical characteristics of the substrate onto which the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected Interacting Domain (SID®) polypeptides are bound.

In yet a further embodiment of the fourth detection method of the invention, the detection step c) consists of bringing into contact the substrate wherein complexes are formed between the targeted polypeptide molecule contained in the sample assayed and the Selected Interacting Domain (SID®) polypeptide or the plurality of Selected Interacting Domain (SID®) polypeptides bound to said support, with a detectable molecule having the ability to bind to such complexes.

A further object of the invention consists of a device or an apparatus for the detection of a polypeptide or a plurality of polypeptides of interest within a sample, wherein said device or apparatus comprises a substrate onto which a Selected Interacting Domain (SID®) polypeptide (or a plurality of Selected Interacting Domain (SID®) polypeptides) is (are) immobilised.

Such a device or apparatus of the invention above may comprise or consist of a suitable substrate onto which the plurality of Selected Interacting Domain (SID®) polypeptides are arranged in an ordered manner, thus forming an area such as described above.

PHARMACEUTICAL COMPOSITIONS CONTAINING A SELECTED INTERACTING DOMAIN (SID®) POLYPEPTIDE.

It results from the method according to which a Selected Interacting Domain (SID®) polypeptide of SEQ ID N°1 to 38 has been selected and characterized that such a Selected Interacting Domain (SID®) polypeptide or a variant thereof is both:

(i) endowed with highly specific binding properties to a (bait) polypeptide of interest;
and

(ii) devoided of the biological activity of the naturally occurring protein from which this Selected Interacting Domain (SID®) polypeptide or a variant thereof is derived.

These original properties of a Selected Interacting Domain (SID®) polypeptide of SEQ ID N°1 to 38 or a variant thereof allow its use for interfering with a naturally occurring interaction between a first protein and a second protein within the cell of an organism by the binding of said Selected Interacting Domain (SID®) polypeptide specifically either to said first polypeptide or said second polypeptide.

The (SID®) polypeptides of the invention or variants thereof are capable of interfering with the *in vivo* protein-protein interactions between HCV proteins or between a HCV protein and a protein from the organism which has been infected with the Hepatitis C virus.

For example the SID® polypeptide of SEQ ID N°2 interferes with the naturally occurring interaction between the core and the NS3 protein HCV. Similarly, the SID® polypeptide of SEQ ID N°17 interferes with the interaction between the NS4A and the NS4B proteins (see table 1).

Thus, another object of the invention consists of a pharmaceutical composition comprising a pharmaceutically effective amount of a Selected Interacting Domain (SID®) polypeptide or a variant thereof.

The invention also relates to a pharmaceutical composition comprising a pharmaceutically effective amount of a nucleic acid comprising a polynucleotide encoding a Selected Interacting Domain (SID®) polypeptide of SEQ ID N°1 to 38 or a variant thereof which polynucleotide is placed under the control of an appropriate regulatory sequence.

Preferred nucleic acids are the nucleotide sequences SEQ ID N°39 to 76.

The invention also pertains to a pharmaceutical composition comprising a pharmaceutically effective amount of a recombinant expression vector comprising a polynucleotide encoding the Selected Interacting Domain (SID®) polypeptide or a variant thereof.

The invention also pertains to a method for preventing or curing a viral infection by a hepatitis C virus in a human or an animal, wherein said method comprises a step of administering to the human or animal body a pharmaceutically effective amount of a Selected Interacting Domain (SID®) polypeptide of SEQ ID N°1 to 38 or a variant thereof which binds to a targeted viral or mammal, typically- human protein.

A pharmaceutical composition as described above, wherein said composition is administered by any route, such as intravenous route, intramuscular route, oral route, or mucosal route with an acceptable physiological carrier and/or adjuvant, also forms part of the invention.

The Selected Interacting Domain (SID®) polypeptide or a variant thereof as a medicament for the prevention and/or treatment of pathologies induced by HCV are the most preferred.

The Selected Interacting Domain (SID®) polypeptides of SEQ ID N°1 to 38 as active ingredients of a pharmaceutical composition will be preferably in a soluble form combined with a pharmaceutically acceptable vehicle.

Such compounds which can be used in a pharmaceutical composition offer a new approach for preventing and/or treating pathologies linked to infection by HCV. Preferably, these compounds will be administered by the systemic route, in particular by the intravenous route, by the intramuscular or intradermal route or by the oral route.

Their modes of administration, optimum dosages and galenic forms can be determined according to the criteria generally taken into account in establishing a treatment suited to a patient, such as for example the age or body weight of the patient, the seriousness of his general condition, the tolerance to treatment and the side effects observed, and the like.

The identified compound can be administered to a mammal, including a human patient, alone or in pharmaceutical compositions where they are mixed with suitable carriers or excipients at therapeutically effective doses to treat disorders associated with prokaryotic micro-organism infection. Techniques for formulation and administration of the compounds of the invention may be found in

" Remington's Pharmaceutical Sciences " Mack Publication Co., Easton, PA, latest edition.

For any Selected Interacting Domain (SID®) polypeptide or any variant thereof used according to the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes or encompasses a concentration point or range shown the desired effect in an *in vitro* system. Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g. for determining the LD50, (the dose lethal to 50% of the test population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD50 and ED50 Compounds which exhibit high therapeutic indices are preferred.

The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50, with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilised. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See, e.g. Fingl et al. 1975, in " The Pharmacological Basis of Therapeutics ", CH.I).

Dosage amount and interval may be adjusted individually to provide plasma levels of the active compound which are sufficient to maintain the modulating effects. Dosages necessary to achieve the modulating effect will depend on individual characteristics and route of administration.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgement of the prescribing physician.

The invention also pertains to a method for preventing or curing a viral in a human or an animal, wherein said method comprises the step of administering to the human or animal body a pharmaceutically effective amount of a nucleic acid comprising a polynucleotide encoding a Selected Interacting Domain (SD®) polypeptide of SEQ ID N°1 to 38, or a variant thereof, and wherein said polynucleotide is placed under the control of a regulatory sequence which is functional in said human or said animal.

Preferred polynucleotides are the nucleic acids of SEQ ID N°39 to 76.

The invention also relates to a method for preventing or curing a viral or in a human or an animal, wherein said method comprises the step of administering to the human or animal body a pharmaceutically effective amount of a recombinant expression vector comprising a polynucleotide encoding a Selected Interacting Domain (SD®) polypeptide which binds to a viral or bacterial protein.

Other characteristics and advantages of the invention appear in the remainder of the description with the examples below, without linking the invention in any manner.

EXAMPLES:

Preparation of a HCV genomic collection.

1.A. Collection preparation and transformation in *Escherichia coli*

1.A.1 Fragmentation of genomic DNA preparation.

The genomic DNA of the infectious HCV strain H77 (Yanagi et al., P.N.A.S. 1997, 94, 8738-43) is fragmented in a nebulizer (GATC) for 2 minutes at a pressure of 2 bars, precipitated and resuspended in water.

The obtained nubilized genomic DNA is successively treated with Mung Bean Nuclease (Biolabs) (30 minutes at 30°C), T4 DNA polymerase (Biolabs) (10 minutes at 37°C) and Klenow enzyme (Pharmacia) (10 minutes at room temperature and 1 hour at 16°C).

DNA is then extracted, precipitated and resuspended in water.

1.A.2. Ligation of linkers to blunt-ended genomic DNA

Oligonucleotide HGX931 (5' end phosphorylated) 1 µg/µl and HGX932 1µg/µl.

Sequence of the oligo HGX931: 5'-GGGCCACGAA-3' (SEQ ID N°151).

Sequence of the oligo HGX932: 5'-TTCGTGGCCCCTG-3'(SEQ ID N°152).

Linkers were preincubated (5 minutes at 95°C, 10 minutes at 68°C, 15 minutes at 42°C) then cooled down at room temperature and ligated with genomic DNA inserts at 16°C overnight.

Linkers were further removed on a separation column (Chromaspin TE 400, Clontech), according to the manufacturer's protocol.

1.A.3. Vector preparation

Plasmid pP6 (see figure 13) was prepared by replacing the Spe1/Xho1 fragment of pGAD3S2X with the double-stranded oligonucleotide:

5'CTAGCCATGGCCGCAGGGGCCGCGGCCGCACTAGTGGGGATCCTTAATTAAAG
GGCCACTGGGGCCCCCCGTACCGGCGTCCCCGGCGCCGGCGTGATCACCCCTA
GGAATTAATTTCCCGGTGACCCCGGGGGAGCT 3' (SEQ ID N°153).

The pP6 vector is successively digested with Sfi1 and BamHI restriction enzymes (Biolabs) for 1 hour at 37°C, extracted, precipitated

and resuspended in water. Digested plasmid vector backbones are purified on a separation column (Chromaspin TE 400, Clontech), according to the manufacturer's protocol.

1.A.4 Ligation between vector and insert of genomic DNA

The prepared vector is ligated overnight at 15°C with the genomic blunt-ended DNA described in section 2 using T4 DNA ligase (Biolabs). The DNA is then precipitated and resuspended in water.

1.A.5. Library transformation in *Escherichia coli.*

Transform DNA from section 1.A.4. into Electromax DH10B electrocompetent ells (Gibco BRL) with Cell Porator apparatus (Gibco BRL). Add 1 ml SOC medium and incubate transformed cells at 37°C for 1 hour. Add 9 ml volume of SOC medium per tube and plate on LB+ampicillin medium. Scrape colonies with liquid LB medium. Aliquot and freeze at -80°C .

The obtained collection of recombinant cell clones is named HGXBHCV1.

1.B. Collection transformation in *Saccharomyces cerevisiae*

The *Saccharomyces cerevisiae* strain (Y187 (MATα Gal4Δ Ga180Δ ade2-101 His3 Leu2-3, -112 Trp1-901 Ura3-52 URA3::UASGAL1-LacZ Met) transformed with the HGXBHCV1 HCV genomic DNA library.

The plasmid DNA contained in *E. coli* are extracted (Qiagen) from aliquoted *E. coli* frozen cells (1.A.5.).

Grow *Saccharomyces cerevisiae* yeast Y187 in YPGlu.

Yeast transformation is performed according to standard protocol (GIEST et al. Yeast, 11, 355-360, 1995) using yeast carrier DNA (Clontech). This experiment leads to 10^4 to 5.10^4 cells/µg DNA. Spread 2.10^4 cells on DO-Leu medium per plates. Aliquot and freeze at -80°C. The obtained collection of recombinant cell clones is named HGXYHCV1.

1.C. Construction of bait plasmids

Plasmid pB5 (see figure 12) is prepared by replacing the NcoI/SalI polylinker fragment with the double-stranded oligonucleotide.

5'CATGGCCGCAGGGGCCGCGGCCGCACTAGTGGGGATCCTTAATTAAAGGGCCA
CTGGGGCCCCCCGGCGTCCCCGGCGCCGGCGTGATCACCCCTAGGAATTAATTT
CCCGGTGACCCCGGGGGAGCT 3'.(SEQ ID N°154).

The linkered genomic DNA described in section 2 is ligated into pB5 that has been digested with Sfi1 restriction enzyme and DNA transformed into competent *E. coli.* Cells are grown and plasmid DNA extracted and sequenced. Those plasmids which code in-frame fusion proteins are used as bait plasmids.

EXAMPLE 2 : Screening the collection with the two-hybrid in yeast system.

2.A. The mating protocol.

We have chosen the mating two-hybrid in yeast system (firstly described by FROMONT-RACINE et al., Nature Genetics, 1997, vol. 16, 277-282, Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens) for its advantages but we could also

screen the HCV collection in classical two-hybrid system as described in Fields et al. or in a yeast reverse two-hybrid system.

The mating procedure allows a direct selection on selective plates because the two fusion proteins are already produced in the parental cells. No replica plating is required. This protocol is written for the use of the library transformed into the Y187 strain.

Before mating, transform *S. cerevisiae* (CG 1945 strain (MATa Ga14-542 Gal180-538 ade2-101 His3*200 Leu2-3, -112 Trp1-901 Ura3-52 Lys2-801 URA::GAL4 17 mers (X3)- CyC1TATA-LacZ LYS2::GAL1UAS-GAL1TATA-HIS3 CYH^R)) according to step 1.B. and spread on DO-Trp medium.

Day 1, morning: preculture

Preculture of Y187 cells carrying the bait plasmid obtained at step 1.C. in 20 ml DO-Trp medium. Grow at 30°C with vigorous agitation.

Day 1, late afternoon: culture

Measure OD_{600nm} of the DO-Trp pre-culture of Y187 cells carrying the bait plasmid preculture. The OD_{600nm} must lie between 0.1 and 0.5 in order to correspond to a linear measurement.
Inoculate 50 ml DO-Trp at OD_{600nm} 0.006/ml, grow overnight at 30°C with vigorous agitation.

Day 2 : mating

<u>medium and plates</u>

1 YPGlu 15 cm plate

50 ml tube with 13 ml DO-Leu-Trp-His

100 ml flask with 5 ml of YPGlu

8 DO-Leu-Trp-His plates

2 DO-Leu plates

2 DO-Trp plates

2 DO-Leu-Trp plates

Measure OD_{600nm} of the DO-Trp culture. It should be around 1.

For the mating, you must use twice as many bait cells as library cells. To get a good mating efficiency, you must collect the cells at 10^8 cells per cm^2.

Estimate the amount of bait culture (in ml) that makes up 30 OD_{600nm} units for the mating with the prey library.

Thaw a vial containing the HGXYHCV1 library slowly on ice. Add the 0.5 ml of the vial to 5 ml YPGlu. Let those cells recover at 30°C, under gentle agitation for 10 minutes.

Mating

Put the 30 OD_{600nm} units of bait culture into a 50 ml flacon tube.

Add the HGXYHCV1 library culture to the bait culture. Centrifuge, discard the supernatant and resuspend in 0.8 ml YPGlu medium.

Distribute the cells onto a YPGlu plate with glass beads. Spread cells by shaking the plates.

Incubate the plate cells-up at 30°C for 4 h 30 min.

Collection of mated cells

Wash and rinse the plate with 6 ml and 7 ml consecutively of DO-Leu-Trp-His.

Perform two parallel serial ten-fold dilutions in 500 µl DO-Leu-Trp-His up to 1/10,000. Spread out 50 µl of each 1/10000 dilution onto DO-Leu and DO-trp plates and 50 µl of each 1/1000 dilution onto DO-Leu-Trp plates.

Spread 3.2 ml of collected cells in 400 µl aliquots on DO-Leu-Trp-His+Tet plates.

DAY 4

Selection of clones able to grow on DO-Leu-Trp-His+Tetracyclin: this medium allows us to isolate diploid clones presenting an interaction.

Count the Trp+Leu+ colonies on control plates and the total number of His+ colonies on the DO-Leu-Trp-His+Tetracyclin plates.

The number of His+ cell clones will define which protocol is to be processed:

Upon 2.10^6 Trp+Leu+ colonies:

- if number of His+cell clones < 95: then process luminometry protocol on all colonies;
- if number of His+ cell clones > 95 and <5000: then process luminometry protocol on 95 colonies;
- if number of His+ cell clones >500: repeat screen using DO-Leu-Trp-His+Tetracyclin plates containing 3-aminotriazol.

2.B The luminometry assay

Grow His+ colonies overnight at 30°C in microtiter plates containing DO-Leu-Trp-His-Tetracyclin medium with shaking. The day after, dilute 15 times overnight culture into a new microtiter plate containing the same medium. Incubate 5 hours at 30°C with shaking. Dilute samples 5 times and read OD_{600nm}. Dilute again to obtain between 10 000 and 75 000 yeast cells/well in 100 µl final volume.

Per well, add 76 µl of One Step Yeast Lysis Buffer (Tropix), 20 µl Sapphirell Enhancer (Tropix), 4 µl Galacton Star (Tropix), incubate 40 minutes at 30°C.

Measure the β-Gal read-out (L) using a Luminometer (Trilux, Wallach).

Calculate value of OD_{600nm}xL and selected interacting preys having highest values.

At this step of the protocol, we have isolated diploid cell clones presenting interaction. The next step is now to identify polypeptides involved in the selected interactions.

EXAMPLE 3: Identification of positive clones

3.A. PCR on yeast colonies

Introduction

PCR amplification of fragments of plasmid DNA directly on yeast colonies is a quick and efficient procedure to identify sequences cloned into this plasmid. It is directly derived from a published protocol (Wang H. et al., Analytical Biochemistry, 237, 145-146, 1996). However, it is not a standardized protocol: in our hands it varies from strain to strain, and is dependent on experimental conditions (number of cells, Taq polymerase source, etc). This protocol should be optimized to specific local conditions.

MATERIALS

- For 1 well, PCR mix composition is:

32.5 µl water,

5 µl 10X PCR buffer (Pharmacia),

1 µl dNTP 10 mM,

0,5 µl Taq polymerase (85µ/µl -Pharmacia),

0,5 µl oligonucleotide ABS1 10 pmole/µl:5'-GCGTTTGGAATCACTACAGG-3',

0,5 µl oligonucleotide ABS2 10 pmole/µl:5'-CACGATGCACGTTGAAGTG-3'.

- 1N NaOH.

Experiment

Grow positive colonies overnight at 30°C on a 96 well cell culture cluster (Costar), containing 150 µl DO-Leu-Trp-His+Tetracyclin with shaking. Resuspend culture and transfer immediately 100 µl on a Thermowell 96 (Costar).

Centrifuge 5 minutes at 4000 rpm at room temperature.

Remove supernatant. Dispense 5 µl NaOH in each well, shake 1 minute.

Place the Thermowell in the thermocycler (GeneAmp 9700, Perkin Elmer) 5 minutes at 99,9°C and then 10 minutes at 4°C.

In each well, add PCR mix, shake well.

Set up the PCR program as followed:

94°C	3 minutes	
94°C	30 seconds	
53°C	1 minute 30 seconds	x 35 cycles
72°C	3 minutes	
72°C	5 minutes	
15°C	∞	

Check the quality, the quantity and the length of the PCR fragment on agarose gel.

The length of the cloned fragment is the estimated length of the PCR fragment minus 300 base pairs that correspond to the amplified flanking plasmid sequences.

3.B Plasmids rescue from yeast by electroporation

Introduction

The previous protocol of PCR on yeast cell may not be successful, in such a case, we rescue plasmids from yeast by electroporation. This experiment allows the recovery of prey plasmids from yeast cells by transformation of *E.coli* with a yeast cellular extract. We can then amplify the prey plasmid and sequence the cloned fragment.

Material

Plasmid rescue

Glass beads 425-600 µm (Sigma)

Phenol/chloroform (1/1) premixed with isoamyl alcohol (Amresco)

Extraction buffer: 2% Triton X100, 1% SDS, 100 mM NaCl, 10 mM TrisHCl pH 8,0, 1 mM EDTA pH 8.0.

Mix ethanol/NH_4Ac: 6 volumes ethanol with 7.5 M NH_4 Acetate, 70% Ethanol and yeast cells in patches on plates.

Electroporation

SOC medium

M9 medium

Selective plates: M9-Leu+Ampicillin

2 mm electroporation cuvettes (Eurogentec)

Experiment

Plasmid rescue

Prepare cell patch on DO-Leu-Trp-His with cell culture of section 2.C.

Scrape the cell of each patch in Eppendorf tube, add 300 µl of glass beads in each tube, then add 200 µl extraction buffer and add 200µl phenol: chloroform:isoamyl alcohol (25:24:1).

Centrifuge tubes 10 minutes at 15000 rpm.

Transfer 180 µl supernatant to a sterile Eppendorf tube and add to each 500 µl ethanol/NH_4Ac, vortex.

Centrifuge tubes 15 minutes, 15000 rmp at 4°C.

Wash pellet with 200 µl 70% ethanol, remove ethanol and dry pellet,

Resuspend pellet in 10 µl water. Store extracts at -20°C.

Electroporation

Material: Electrocompetent MC1066 cells prepared according to standard protocols (Maniatis).

Add 1 µl of yeast plasmid DNA-extract to pre-chilled Eppendorf tube, and keep on ice.

Mix 1 µl plasmid yeast DNA-extract sample, add 20 µl electrocompetent cells and transfer in a cold electroporation cuvette.

Set the Biorad electroporator on 200 ohms resistance, 25 µF capacity; 2.5 kV. Place cuvette in the cuvette holder and electroporate.

Add 1 ml SOC into the cuvette and transfer the cell-mix into sterile Eppendorf tube.

Let cells recover for 30 minutes at 37°C, spin the cells down 1 minute, 4000x g and pour off supernatant. Keep about 100 µl medium and use it

to resuspend the cells and spread them on selective plates (e.g. M9-Leu plates).

Incubate plates for 36 hours at 37°C.

Grow one colony and extract plasmids. Check presence and size of insert through enzymatic digestion and agarose gel. Sequence insert.

EXAMPLE 4: Protein-protein interaction.

For each bait, the previously protocol leads to the identification of prey polynucleotide sequences. Using a suitable software program (eg Blastwun, available on the Internel site of the University of Washington: http:/bioweb.pasteur.fr/seqanal/interfaces/blastwu.html) the region of the HCV genome is encoded by the prey fragment may be determined and whether the fusion proteins encoded are in the same open reading frame of translation as the HCV polyprotein or not.

EXAMPLE 5 : Identification of SID®

The presence of contiguous polypeptides in the HCV genome and the high complexity of the prey library used prevents the determination of SID®s by previous means since prey fragments can overlap multiple polypeptides. The high complexity of the prey library used relative to the small genome size also prevented such a simple analysis since prey fragments can overlap multiple interacting domains. It was also necessary to overcome the problems caused by protein preys encoded by out-of-frame fusions of regions of the HCV genome.

In order to determine the SID®s for a particular bait protein, it was therefore necessary to devise a suitable algorithm which would take into account all these problems:

5.1. The prey fragments are initially sorted according to which reading frame of the polypeptide sequence they correspond to. This enables the separation of physiologically relevant prey protein from out-of-frame fusions which bind in the two-hybrid assay.

5.2. Each prey fragment is compared pairwise with other prey fragments and two fragments are clustered together if they overlap by more than 30% of their lengths (see fig. 8). Further fragments are assigned to the cluster if, and only if, overlap all the fragments in the cluster by more than 30% of their length.

5.3 For each cluster of fragments thus produced, a pre-SID is defined as the intersection of all the fragments present in the cluster defined in 5.2 (figure 9).

5.4. The pre-SIDs defined in 5.3 are then analysed pairwise and if the region of intersection between two pre-SIDs is greater than 30 bp then a SID® is defined as this region of intersection. If the non-intersecting region of a pre-SID is of more than 30 bp in length and this non-intersecting region represents more than 30% of the length of one of the fragments that comprises this region, then this non-intersecting region is also defined as a SID®s (figure 10).

5.5 The number of fragments contributing to each SID defined in 5.4 is counted. In the case of overlapping SIDs®, the SID® which contains the most fragments is identified, and all the fragments which contribute to this SID® are removed from overlapping SIDs®. The inspection of the fragments which remain in these overlapping SIDs® determines the final sequence of the SID® (figure 11).

TABLE 1

Summary of the protein-proptein interactions between the SID polypeptides of the invention and H77 strain HCV polypeptides

Bait	SEQ ID N°(1)	begin(2)	end(2)	SEQ ID N°(3)	SID	SEQ ID N° (4)	begin (2)	end (2)	SEQ ID N° (5)
Core(87%)	114	302	614	77	Core(100%)	39	446	600	1
Core(100%)	115	342	683	78	NS3(100%)	40	4814	4922	2
Core(100%)	115	342	683	78	Core(100%)	41	380	616	3
E1(100%)	116	995	1342	79	E2(100%)	42	1871	1987	4
E1(4%)/E2(95%)	117	1478	1756	80	NS3(100%)	43	4787	5242	5
E2(100%)	118	1745	2278	81	E2(100%)	44	1871	1958	6
E2(100%)	119	1799	2090	82	E2(100%)	45	1808	1890	8
NS2(12%)/NS3(87%)	120	3312	4150	83	NS4A(59%)/NS4B(40%)	46	5375	5542	9
NS3(100%)	121	3767	4244	84	NS3(100%)	47	4676	4801	10
NS3(100%)	122	3779	4571	85	NS3(100%)	48	4856	4945	11
NS3(100%)	123	3974	4559	86	NS3(100%)	49	4817	4903	12
NS3(100%)	124	4238	4857	87	NS5B(100%)	50	7979	8109	13
NS3(100%)	125	4298	4859	88	NS3(100%)	51	4031	4118	14
NS3(100%)	126	4691	5168	89	E2(100%)	52	1784	1888	15
NS3(100%)	127	4838	5230	90	E2(100%)	53	1871	1968	16
NS3(1%)/NS4A(98%)	128	5310	5467	91	NS4B(100%)	54	5918	6154	17
NS4A(100%)	129	5342	5400	92	NS3(100%)	55	3512	3956	18
NS4B(86%)/NS5A(13%)	130	5717	6344	93	NS4B(53%)/NS5A(46%)	56	6197	6310	19
NS4B(70%)/NS5A(29%)	131	5819	6444	94	E2(100%)	57	1844	1933	20
NS4B(55%)/NS5A(44%)	132	5882	6562	95	NS5B(100%)	58	9083	9222	21
NS4B(82%)/NS5A(17%)	133	5897	6335	96	NS4B(100%)	59	5819	6080	22
NS4B(100%)	134	6011	6177	97	E2(100%)	60	1823	1955	23

TABLE 1 (continued)

Bait	SEQ ID N°(1)	begin(2)	end(2)	SEQ ID N°(3)	SID	SEQ ID N° (4)	begin (2)	end (2)	SEQ ID N° (5)
NS4B(30%)/NS5A(69%)	135	6107	6605	98	NS4B(100%)	61	5879	6072	24
NS4B(12%)/NS5A(87%)	136	6141	7069	99	E2(100%)	62	1784	1875	25
NS4B(8%)/NS5A(91%)	137	6182	7034	100	E1(100%)	63	1226	1458	26
NS4B(9%)/NS5A(90%)	138	6188	6939	101	NS4B(70%)/ NS5A(28%)	64	6176	6291	27
NS5A(100%)	139	6317	6576	102	NS3(100%)	65	4784	4928	28
NS5A(100%)	140	6440	6727	103	NS5A(100%)	66	6557	6721	29
NS5A(100%)	141	7019	7249	104	NS3(100%)	67	4451	4790	30
NS5A(100%)	142	7274	7549	105	NS4B(100%)	68	6029	6194	31
NS5B(100%)	143	7613	8027	106	NS5B(100%)	69	8354	8665	32
NS5B(100%)	144	7838	8743	107	NS5B(100%)	70	7769	8011	33
NS5B(100%)	145	7856	8458	108	NS3(100%)	71	4715	4901	34
NS5B(100%)	146	7976	8759	109	NS5B(100%)	72	7775	8011	35
NS5B(100%)	147	8564	8948	110	E2(100%)	73	1805	1887	36
NS5B(100%)	148	8708	8978	111	E2(100%)	74	1751	1865	37
NS5B(100%)	149	8996	9220	112	NS4B(57%)/ NS5A(41%)	75	6194	6303	38
NS5B(100%)	150	9032	9226	113	NS4B(63%)/ NS5A(35%)	76	6206	6286	39

(1) Nucleic acid sequence encoding the polypeptide from the H77 strain of HCV which binds to the SID polypeptide (4) described in the same line.

(2) 5'-end and 3'-end nucleotide positions of the sequence SEQ ID (1) in reference to the nomenclature disclosed by Yanagi et al. (1997)

(3) Aminoacid sequence of the polypeptide from the H77 strain of HCV which binds to the SID polypeptide (4) described in the same line.

(4) Nucleic acid sequence encoding the SID polypeptide which binds to the polypeptide of the aminoacid sequence (3) described in the same line.

(5) Aminoacid sequence of the SID polypeptide which binds to the polypeptide of the aminoacid sequence (3) described in the same line.

REFERENCES

- Aguzzi F et al. , Estratto Dal. Boll. 1 st Sieroter; Milanese, 1977, vol.56: 212-216.
- Brigham KL et al., 1993, Am. J. Respir Cell Mol. Biol.8(2):209-213
- BARTENSCHLAGER R. et al., 1995, J. Virol., 69 (12): 7519-7528.
- Curiel et al. Gene Transfer to Respiratory Epithelial Cells via the Receptor Mediated Endocytosis Pathway, Am. J. Respir. Cell Mol. Biol. 6 (1992) 247-252
- Curiel et al. Adenovirus Enhancement of Transferrin-Polylysine-Mediated Gene Delivery, Proc. Natl. Acad. Sci. 88 (1991) 8850-8854
- Chalfie et al., (1994), Science, vol. 263: 802-805.
- Chen F-T.A. et al. , 1991, Clin. Chem., vol.77: 14-19.
- Davis et al., (1995), Development Biology, vol.170: 726-729
- Delagrave et al., (1995), Biotechnology, vol.13: 151-154
- DI MARCO et al., 2000, The Journal of Biological Chemistry, vol.275 (10):7152-7157
- DUBUISSON J., 1994, J.Viral. vol.68:6147-6160
- Drumm, M. L. et al., Cell 62:1227-1233 (1990).
- Fromont-Racine M et al. (1997), Nature Genetics, vol.16 (3):277-282.
- FLAJOLET M. et al., 2000 Gene, vol.242:369-379.
- EDWARDS and LEARTHERBARROW, 1997, Analytical Biochemistry, 246:1-6.
- Goding et al. J.W., (1986), In: Monoclonal antibodies: Principles and practice-production and application of monoclonal antibodies in cell biology, biochemistry and immunology, Acad. Press, London, pages 255-280.
-
- GALLINARY P et al., 1999, Biochemistry, vol.38:5620-5632
- GRAKOUI A., 1993, J.Viral., vol.67:1385-1395.

- HOUGHTON, M (1996), Hepatitis C virus, fields editors.
- HIGIKATA M., 1993, J.Viral., vol.67:4665-4675
- Heim et al., (1994), Proc. Natl. Acad. Sci., volume 91: 12501-12.504.
- Hu and Cheng, (1995), Febs. Letters, vol.369: 331-334.
- Ichinose N et al;, (1991), In: Fluorometric analysis in biomedical chemistry, vol.10, page 110, Chemical analysis, Winefordner JD et al. Eds., John Wiley and Sons, New York.
- KARIMOVA et al., 1998, Proc. Natl. Acad. Sci., USA, 95:5752-5756.
- Keegan et al. (1986), Science, vol.231 (4739): 699-704.
- KOCH Y, 1977, Biochem. Biophys. Res. Commun, vol.74:488-491
- Kohler and Milstein, 1975, Nature, 256: 495
- Kozbor et al., 1983, Hybridoma, 2(1):7-16.
- KEEGAN et al., 1986, Science, Vol.231:689-407
- Kaether and Gerdes, (1995), Febs. Letters, vol.369:267-271.
- Leger et al., 1997, Hum. Antibodies, 8(1):3-16
- Martineau et al., 1998, J. Mol. Biol., 280(1): 117-127
- Muzyczka, N., Curr. Top. Micro. Immuno. 158:97-129 (1992)
- Merlini G et al. , 1983, J. Clin. Chem. Biochem. , vol.21: 841-844,
- Maggio ET, " Enzyme-immuno assay ", 1980, CRC Press Incorporated , Boca Raton, Fla.
- MA and PATSHNE, 1987, Cell, vol. 48: 847-853
- MIN et al., 1999, virus genes, vol.19 (1):33-43
- Nielsen et al., 1991, J. Chromatogr., vol.539: 177
- PATEL J. et al. 1999, Journal of General Virology, vol.80:1681-1690.
- Pontiroli et al., 1987, Diabet. Metab., vol.13:441-443.
- Ridder et al., 1995, Biotechnology (NY), 13(3):255-260
- Reinmann et al., 1997, AIDS Res. Hum. Retroviruses, 13(II): 933-943
- Rosenfeld, M. A. et al., Cell 68:143-155 (1992)

- Rizzuto et al. , (1995), Current Biology, vol.5: 635-142.
- ROUGEOT C et al., 1994, Eur. J. Biochem., vol.219(3):765-773
- Shattil SJ et al., (1987) , Blood, vol.70: 307.
- Shattil et al. SJ(1985), J. Biol. Chem., vol.260:11.107.
- Smith *et al.*, 1988, Gene 67:31-40.
- Schofield, Brit. Microencapsulated. Bull., 51(1):56-71 (1995)
  Behr, Bioconjugate Chem., 5, 382-389 (1994)
- SZABO A. et al., 1995, Curr. Opin. Struct. Biol., 5(5): 699-705.
- Trubetskoy, V. S. et al., Biochem. Biophys. Acta 1131:311-313 (1993))
- UTKIEWICZ NJ et al., 2000, vol. 267: 278-282
- URBANI A et al., 1999, Biochemistry, vol.38:5206-5215
- Wu et al., 1992, J. Biol. Chem. 267:963-967.
- Wu and Wu, 1988, J. Biol. Chem. 263:14621-14624.
- Wilson, J. M. et al., 1992, Endocrinology, 130(5):2947-2954
- White Wa et al. , 1986, Biochem. Clin. vol.10:571-574.
- Yanagi et al., Proc. Nat. Acad. Sci USA, 1997, 94 :8738-8743